

# Computergestützte Strukturbestimmung biochemischer Komplexe durch einen *Fuzzy Logic*-basierten Algorithmus

Vom Fachbereich Chemie  
der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades eines

Doktor - Ingenieurs (Dr.-Ing.)

genehmigte  
*Dissertation*

vorgelegt von

Dipl.-Ing. Thomas Eckart Exner

aus Stuttgart

Berichterstatter:

Mitberichterstatter:

Tag der Einreichung:

Tag der mündlichen Prüfung:

Prof. Dr. J. Brickmann

Prof. Dr. A. Kerber

Prof. Dr. M. Martin

17. Februar 2000

17. April 2000

Darmstadt 2000

*„Corpora non agunt nisi fixata“*

Paul Ehrlich, 1913

*Meiner Familie gewidmet*

Die vorliegende Arbeit wurde im Fachbereich Chemie, Fachgebiet Physikalische Chemie I, auf Anregung und unter Anleitung von Herrn Prof. Dr. J. Brickmann in der Zeit von November 1996 bis Januar 2000 durchgeführt.

Mein Dank gilt an dieser Stelle

- Herrn Prof. Dr. Jürgen Brickmann für die interessante Themenstellung, die stetige Unterstützung und die mir gewährte Entfaltungsfreiheit,
- Herrn Dr. Hans-Jürgen Bär für die fachlichen Anregungen und die sorgfältige Korrektur des Manuskripts,
- Herrn Dipl.-Ing. Matthias Keil für die Unterstützung bei wissenschaftlichen und computertechnischen Fragestellungen,
- den Herren Dr. Gerd Moeckel, Dipl.-Ing. Robert Jäger, Dipl.-Ing. Thorsten Borosch sowie meinem Vater Prof. Dr. Hans Eckart Exner für interessante und weiterführende Diskussionen,
- und allen Mitgliedern des Arbeitskreises für die sowohl fachlich anregende als auch freundschaftliche Atmosphäre, die die Arbeit sehr erleichterte;
- sowie meiner Freundin Carole Guérin für das offene Ohr, das sie immer für meine Probleme hatte.
- Darüber hinaus danke ich den vielen, hier nicht namentlich genannten Personen, die auf die eine oder andere Art zum Gelingen dieser Arbeit beigetragen haben.

## Inhaltsverzeichnis

<b>1. EINLEITUNG.....</b>	<b>1</b>
<b>2. THEORETISCHE UND MATHEMATISCHE GRUNDLAGEN.....</b>	<b>4</b>
2.1 Molekulare Erkennung in biochemischen Komplexen.....	4
2.2 <i>Docking</i> -Algorithmen .....	7
2.3 <i>Fuzzy Set Theory</i> und <i>Fuzzy Logic</i> .....	9
2.3.1 Definition von <i>Fuzzy</i> -Mengen .....	10
2.3.2 Linguistische Variable .....	11
2.3.3 Verknüpfungsoperatoren von <i>Fuzzy</i> -Mengen.....	12
2.4 Genetische Algorithmen .....	14
2.4.1 Selektion .....	14
2.4.2 Crossover .....	15
2.4.3 Mutation.....	15
2.4.4 Weitere Operatoren genetischer Algorithmen .....	15
<b>3. TOPOGRAPHISCHE UND PHYSIKOCHEMISCHE EIGENSCHAFTEN</b>	
<b>    MOLEKULARER OBERFLÄCHEN .....</b>	<b>17</b>
3.1 Molekulare Oberflächen .....	17
3.2 Topographische Analyse der molekularen Oberfläche .....	18
3.2.1 Kanonische und globale Krümmungen.....	19
3.2.2 Beschreibung der Form und Neigung molekularer Oberflächen .....	21
3.3 Elektrostatisches Potential .....	24
3.4 Lokale Lipophilie/Hydrophilie .....	26
3.5 Fähigkeit zur Ausbildung von Wasserstoffbrückenbindungen.....	27
3.6 <i>Fuzzifizierung</i> der topographischen und physikochemischen Eigenschaften .....	28
<b>4. ALGORITHMISCHE KONZEPTE ZUR LÖSUNG DES <i>DOCKING</i>-PROBLEMS..</b>	<b>31</b>
4.1 Segmentierung der molekularen Oberflächen.....	32
4.2 Vergleich der Domänen der molekularen Oberflächen .....	36
4.2.1 Bestimmung der Eigenschaften einer Domäne.....	36
4.2.2 Vergleich der Domäneneigenschaften .....	37
4.2.3 <i>Geometric Hashing</i> -Algorithmus .....	39
4.2.4 Test auf sterische Überlappung.....	42

4.3	Clusteralgorithmus zur Bestimmung mittlerer Orientierungen aus ähnlichen Komplexstrukturen .....	43
4.4	Zielfunktion für die Optimierung der Komplexstrukturen .....	46
4.5	<i>Downhill Simplex</i> -Minimierer .....	50
4.6	Genetischer Algorithmus zur flexiblen Optimierung von Komplexstrukturen .....	53
<b>5.</b>	<b>ERGEBNISSE UND DISKUSSION .....</b>	<b>56</b>
5.1	Aufarbeitung der kristallographischen Strukturdaten .....	58
5.2	Segmentierung der molekularen Oberflächen.....	59
5.3	Starres <i>Docking</i> .....	62
5.3.1	Vorhersage der Komplexstrukturen niedermolekularer Liganden.....	62
5.3.2	Komplex der Dihydrofolat-Reduktase mit Methotrexat (4dfr).....	68
5.3.3	Beschreibung weiterer Komplexstrukturen .....	70
5.3.4	Vorhersage für Enzym-Ligand-Komplexe mit inhibierenden Proteinen, Protein- Dimere und Antikörper-Antigen-Komplexe.....	72
5.3.5	Bewertung des vorgeschlagenen <i>Docking</i> -Algorithmus.....	74
5.4	Flexibles <i>Docking</i> von Methotrexat in Dihydrofolat-Reduktase .....	75
<b>6.</b>	<b>ZUSAMMENFASSUNG .....</b>	<b>79</b>
<b>LITERATUR .....</b>		<b>82</b>
<b>ANHANG 1: DOMÄNEN DER ENZYM-INHIBITOR-KOMPLEXE MIT NIEDER- MOLEKULAREN INHIBITOREN.....</b>		<b>95</b>
<b>ANHANG 2: DOMÄNEN DER ENZYM-INHIBITOR-KOMPLEXE MIT INHIBIERENDEN PROTEINEN .....</b>		<b>99</b>
<b>ANHANG 3: DOMÄNEN DER PROTEIN-DIMERE .....</b>		<b>102</b>
<b>ANHANG 4: DOMÄNEN DER ANTIKÖRPER-ANTIGEN-KOMPLEXE .....</b>		<b>103</b>
<b>ANHANG 5: DOCKING-ERGEBNISSE FÜR ENZYM-LIGAND-KOMPLEXE MIT INHIBIERENDEN PROTEINEN, PROTEIN-DIMERE UND ANTI-KÖRPER-ANTIGEN-KOMPLEXE .....</b>		<b>104</b>
<b>ANHANG 6: ERGEBNISSE DES FLEXIBLEN DOCKINGS VON METHOTREXAT IN DIHYDROFOLAT-REDUKTASE.....</b>		<b>108</b>

## Kapitel 1

### Einleitung

Die Kosten für die Entwicklung und Einführung eines neuen Arzneimittels sind in den letzten Jahrzehnten überproportional auf etwa 300-600 Millionen DM gestiegen. Diese steigenden Kosten der Forschung und Entwicklung, ein bereits bestehender hoher Therapiestandard in vielen Indikationsgebieten sowie ein verstärktes Sicherheitsbewußtsein und damit zunehmende Anforderungen von Seiten des Gesetzgebers haben dazu geführt, daß die Zahl der in die Therapie neu eingeführten Wirkstoffe stetig abgenommen hat [1]. Für deren schnellere und kostengünstigere Herstellung wird deswegen die Einführung von neuen Methoden für den rationalen Entwurf, wie z.B. das struktur- und computergestützte Design neuer Medikamente, immer wichtiger.

Beim rationalen Entwurf wird versucht, den Wirkmechanismus zu analysieren und auf der Basis dieses Wissens gezielt Vorschläge für eine molekulare Leitstruktur zu entwickeln. Um als Arzneimittel wirken zu können, muß diese Leitstruktur im Körper an ein ganz bestimmtes Zielmolekül binden. Deswegen stellt der erste Schritt der Medikamentenentwicklung die Identifizierung und Charakterisierung des für die spezielle biochemische Reaktion verantwortlichen Rezeptors dar. Ist dieser bekannt, kann zur Ermittlung und Optimierung der Leitstruktur auf computergestützte Methoden zurückgegriffen werden. Dadurch kann das teure und zeitaufwendige Testen der Affinität einer sehr großen Menge von Verbindungen (*Screening*) vermieden werden

Zur Bestimmung der Affinität eines Rezeptors mit Hilfe des Computers können verschiedene Ansätze verwendet werden. Ausgehend von den physikochemischen Eigenschaften der beteiligten Partner können durch die quantitative Analyse von Struktur-Wirkungsbeziehungen (*Quantitative Structure Activity Relationship, QSAR*) empirische Vorhersagen von Komplexbindungsenergien gemacht werden, und daraufhin kann eine Optimierung der Leitstruktur erfolgen. Ist zusätzlich die dreidimensionale Struktur des Rezeptors aus Röntgenstrukturanalysen oder NMR-Experimenten bekannt, ist durch die Charakterisierung des aktiven Zentrums und der Vorhersage möglicher Orientierungen, sogenannter Bindungsmodi, der wechselwirkenden Moleküle eine direkte Berechnung der Bindungsenergie möglich. Bei der Beschreibung dieser molekularen Erkennung kommen eine Vielzahl unterschiedlicher Faktoren (energetisch, entropisch usw.) ins Spiel. Für eine schnelle Ermittlung der Komplexstrukturen, im weiteren *Docking* genannt, ist deswegen eine effektive Umsetzung dieser Faktoren in ein mathematisches Modell von entscheidender Bedeutung.

Für eine erfolgreiche Komplexbildung müssen alle Bindungsregionen des Rezeptors in ihrer Form stark komplementär zur Bindungsregion des Liganden sein. Die wohl effektivste Strategie für die Suche nach Formkomplementarität ist wahrscheinlich die direkte visuelle Inspektion durch das menschliche Auge [2], da es für den Menschen relativ leicht herauszufinden ist, ob ein Objekt möglicherweise zu der Oberfläche paßt oder nicht. Durch moderne Methoden der interaktiven Computergraphik ist es möglich, mit mathematisch definierten Objekten reale Szenarien nachzubilden. So kann die molekulare Erkennung in diese Modellwelt übertragen werden, indem durch den Vergleich von mathematisch definierten Moleküloberflächen Komplexstrukturen ermittelt werden. Die visuelle Inspektion hat jedoch ihre Grenzen, wenn eine sehr große Anzahl von Objekten untersucht werden soll. Durch die Umsetzung der menschliche Fähigkeit zur Formerkennung in Algorithmen kann dagegen die Leistungsfähigkeit der modernen Computertechnologie zum Einsatz kommen. Deswegen werden in der hier vorgestellten Arbeit Ansätze entwickelt, die zur Beschreibung der molekularen Erkennung die menschliche Fähigkeit der Formerkennung und Problemoptimierung in mathematische Algorithmen umsetzen. Es wird das Ziel verfolgt, mit Hilfe dieser Algorithmen die aktive Struktur biochemischer Komplexe ohne Vorkenntnis über die Bindungsregion zu bestimmen. Dazu soll zunächst eine Auswahl möglicher Komplexstrukturen bestimmt werden, aus der anschließend der biologisch aktive Komplex mit Hilfe einer Bewertungsfunktion extrahiert wird.

Zur formalen Beschreibung der menschlichen Denkweisen hat sich die *Fuzzy Set Theory* in vielen Einsatzgebieten als erfolgreich erwiesen [3,4]. Sie wurde deswegen in der hier vorgestellten Arbeit zur Beschreibung des molekularen Erkennungsprozesses verwendet. Das Konzept der *Fuzzy Set Theory* bzw. *Fuzzy Logic* (unscharfe Logik) wurde als Erweiterung der klassischen Mengenlehre bzw. der Boole'schen Logik eingeführt [5]. Durch die dadurch mögliche unscharfe Beschreibung der molekularen Eigenschaften konnte in Kombination mit einer effektiven Energieoptimierung ein *Docking*-Algorithmus zur Bestimmung von Komplexstrukturen biochemischer Moleküle entwickelt werden. Dieser wurde an bekannten biochemischen Komplexen getestet. Dabei wurden die Konformationen aus den kristallographisch ermittelten Komplexdaten verwendet. Anschließend wurde der Algorithmus auf *Docking*-Experimente, bei denen die Flexibilität des Liganden berücksichtigt wird, erweitert.

Die Arbeit ist wie folgt gegliedert:

In Kapitel 2 werden die chemischen und mathematischen Grundlagen beschrieben. Dabei soll auf die verschiedenen Arten von biochemischen Komplexen, die durch diese hervorgerufenen physiologischen Wirkungen sowie auf die bereits in der Literatur vorgestellten *Docking*-Ansätze eingegangen werden. Als mathematische Grundlagen werden die *Fuzzy Set Theory* sowie die Funktionsoptimierung mit Hilfe genetischer Algorithmen erläutert. Das dritte

Kapitel befaßt sich mit der Beschreibung des in dieser Arbeit verwendeten Modells einer molekularen Oberfläche und der Berechnung der topographischen und physikochemischen Eigenschaften am Ort dieser Oberflächen. Kapitel 4 beschreibt die Details der algorithmischen Ansätze zur Lösung des *Docking*-Problems. In Kapitel 5 folgen dann die Zusammenstellung und die Diskussion der Ergebnisse. Dabei werden zuerst die Ergebnisse des *Dockings* von 35 ausgewählten Komplexen unter Verwendung starrer Molekülstrukturen zusammengefaßt. Anschließend folgt an einem Komplex exemplarisch die Erweiterung des Ansatzes durch die Berücksichtigung der Flexibilität. Die Arbeit wird durch eine kurze Zusammenfassung des Algorithmus und der erhaltenen Ergebnisse in Kapitel 6 abgeschlossen.



## Kapitel 2

### Theoretische und mathematische Grundlagen

In diesem Kapitel werden die für die Entwicklung des hier vorgestellten *Docking*-Algorithmus benötigten biochemischen und mathematischen Grundlagen beschrieben. Kapitel 2.1 gibt eine kurze Einführung in die molekularen Erkennungsprozesse und die für die algorithmische Beschreibung benötigten biochemischen Grundlagen. In Kapitel 2.2 folgt eine Zusammenstellung der in der Literatur beschriebenen Ansätze zur Lösung des *Docking*-Problems. In den Kapiteln 2.3 und 2.4 wird kurz auf die verwendeten mathematischen Methoden der *Fuzzy Set Theory* und der genetischen Algorithmen eingegangen.

#### 2.1 Molekulare Erkennung in biochemischen Komplexen

Die spezifische Erkennung zweier Moleküle und die anschließende Ausbildung eines Komplexes spielen eine Schlüsselrolle in vielen biochemischen Prozessen. Die Vielzahl verschiedener Aufgaben wird durch unterschiedliche Arten von Komplexen gelöst [6,7].

Nahezu alle biochemische Reaktionen werden durch Enzyme katalysiert. Dabei bindet das Substrat an eine spezifische Region des Enzyms, die als aktives Zentrum bezeichnet wird. Trotz der starken Unterschiede in Struktur, Spezifität und Art der Katalyse können einige allgemeine Aussagen über enzymatisch katalysierte Reaktionen gemacht werden:

- Die Reaktionsgeschwindigkeiten sind in der Regel  $10^6$  bis  $10^{12}$  mal größer als die entsprechenden nichtkatalysierten Reaktionen.
- Es werden nur relativ milde Reaktionsbedingungen benötigt: Temperaturen deutlich unter  $100^\circ\text{C}$ , Atmosphärendruck und fast neutrale pH-Werte.
- Enzyme können eine große Spezifität im Hinblick auf die Identität der Substrate (Edukte) und Produkte besitzen.
- Die katalytische Aktivität kann durch Substanzen, die nicht Substrate sind, reguliert werden.

Das aktive Zentrum besteht aus einer relativ kleinen dreidimensionalen Einheit der Enzymstruktur. Die katalytische Aktivität und die Bindungsspezifität werden durch die spezielle Anordnung der Atome und deren Wechselwirkungen bestimmt. Das Substrat wird während der Reaktion in einer genau definierten Lage an das Enzym gebunden. Die katalytische Wirkung des Enzyms kann dann durch verschiedene Mechanismen, wie z.B. durch Stabilisierung des Übergangszustandes oder durch Protonen- oder Elektronentransfer, erfolgen. Wird das aktive Zentrum durch ein anderes Molekül (Inhibitor) blockiert, kann das Enzym seine katalytische Wirkung nicht mehr entfalten, und die durch diese Reaktion

hervorgerufene physiologische Wirkung wird unterdrückt. Deswegen eignen sich die Enzyme hervorragend als Ansatzpunkt für die Medikamentenentwicklung. So müssen Inhibitoren entwickelt werden, die eine große Komplexbildungskonstante aufweisen und dadurch eine dauerhafte Blockierung des aktiven Zentrums ermöglichen. Durch die Identifizierung und Charakterisierung der Bindungsmodi können theoretische Vorhersagen gemacht werden, die zu Medikamenten mit größerer Wirkung und Spezifität führen können. Neben dieser kompetitiven Hemmung spielen zwei weitere Mechanismen bei der Inhibierung der Enzymfunktion eine Rolle. Bei der nichtkompetitiven sowie der unkompetitiven Hemmung bindet der Inhibitor an eine andere Bindungsregion als das Substrat. Die beiden Mechanismen unterscheiden sich dadurch, daß im Fall der nichtkompetitiven Hemmung die Substratbindung durch strukturelle Änderungen der Bindungsregion verhindert wird. Bei der unkompetitiven Hemmung bindet zuerst das Substrat an das Enzym. Die anschließende Bindung des Inhibitors bewirkt dann eine Veränderung der Konformation des aktiven Zentrums, wodurch die Reaktion des Substrats verhindert wird.

Viele Proteine bestehen nicht nur aus einer Polypeptid-Kette, sondern sind aus mehreren Untereinheiten aufgebaut. Dabei kann es sich bei den Untereinheiten um identische oder nichtidentische Polypeptid-Ketten handeln. Proteine mit identischen Untereinheiten werden als Oligomere, die mit nichtidentischen Untereinheiten als Protomere bezeichnet. Der Aufbau dieser Quartärstrukturen bildet meistens erst die Grundlage für die biochemische Aktivität des Proteins sowie die strukturelle Basis für die Regulation dieser Aktivität. Die Kontaktregionen zwischen zwei Untereinheiten ähneln dabei dem Inneren eines Proteins. Sie enthalten dicht gepackte, unpolare Seitenketten, Wasserstoffbrückenbindungen sowie in einigen Fällen auch Disulfid-Bindungen zwischen verschiedenen Ketten.

Die humorale Immunität beruht auf der Bildung einer Vielzahl ähnlicher Proteine, die als Antikörper oder Immunglobuline bezeichnet werden [6]. Dabei bildet sich zwischen dem Antigen und einem Antikörper, der eine speziell auf dieses Antigen angepaßte Bindungsregion aufweist, ein Komplex. Die Antigen-Bindungsstelle liegt an der Spitze der beiden FAB-Regionen in der Spalte zwischen der variablen Einheit der leichten Kette  $V_L$  und der schweren Kette  $V_H$ . Die Größe und Form dieser Spalte hängt von der Aminosäuresequenz der  $V_L$ - und  $V_H$ -Einheit ab. Daraus ergibt sich die Spezifität des Antikörpers für ein bestimmtes Antigen.

Die nicht kovalenten Kräfte, die eine Bindung zwischen den Molekülen vermitteln, sind in allen diesen Komplexen in ihrer Art identisch: van-der-Waals-Bindungen, elektrostatische und hydrophobe Wechselwirkungen sowie Wasserstoffbrückenbindungen. Deswegen ist eine gemeinsame, theoretische Behandlung aller dieser Komplexe möglich. Im weiteren sollen die Komplexpartner unabhängig vom Komplex typ wie folgt benannt werden: Das jeweils größere

Molekül wird als Rezeptor bezeichnet, das kleinere als Ligand. Bei Protein-Dimeren wird jeweils die erste Polypeptid-Kette als Rezeptor angesehen. Die zweite Polypeptid-Kette entspricht dann dem Liganden. Die Ausbildung des Komplexes geschieht bei allen Komplexarten in einer ausgeprägten Region, im weiteren als Bindungsregion oder *Docking-Region* bezeichnet. Diese unterscheiden sich in den einzelnen Komplexarten, aber auch in verschiedenen Komplexen derselben Art, stark in Form und Größe. Für eine erfolgreiche Komplexbildung müssen alle Bindungsregionen des Rezeptors in ihrer Form stark komplementär zur Bindungsregion des Liganden sein (geometrische Komplementarität). Die Aminosäure-Reste des Rezeptors sind zusätzlich so angeordnet, daß sie spezifisch mit der Bindungsregion des Liganden wechselwirken können. Daraus ergibt sich die Komplementarität der elektrostatischen und hydrophoben Eigenschaften; außerdem wird die Ausbildung von spezifischen Wasserstoffbrücken möglich.

Fischer [8] entwickelte die Vorstellung, daß die Bindungsstelle des Rezeptors bereits in Abwesenheit des Liganden vorgeformt vorliegt. Nach dieser Schlüssel-Schloß-Hypothese ergibt sich die Spezifität des Rezeptors (Schloß) für einen Liganden (Schlüssel) aus der geometrisch komplementären Form (Abb. 2.1). Als Weiterentwicklung beschreibt die *Induced Fit*-Hypothese, daß sich die Bindungsstelle erst ausbildet, wenn der Ligand an den Rezeptor bindet (Abb. 2.2). Durch Röntgenstrukturanalysen konnte gezeigt werden, daß bei den meisten Rezeptoren die Bindungsstelle bereits vorgeformt ist, es aber dennoch während der Komplexbildung zu kleinen Konformationsänderungen im Sinne des *Induced Fit* kommt [6,7]. Sind diese Verhältnisse gegeben, sollte eine theoretische Bestimmung der Komplexstruktur möglich sein, indem die Komplementarität der Oberflächen der Komplexpartner ermittelt und die Flexibilität der Moleküle nur im Bereich dieser komplementären Oberflächenbereiche berücksichtigt wird.

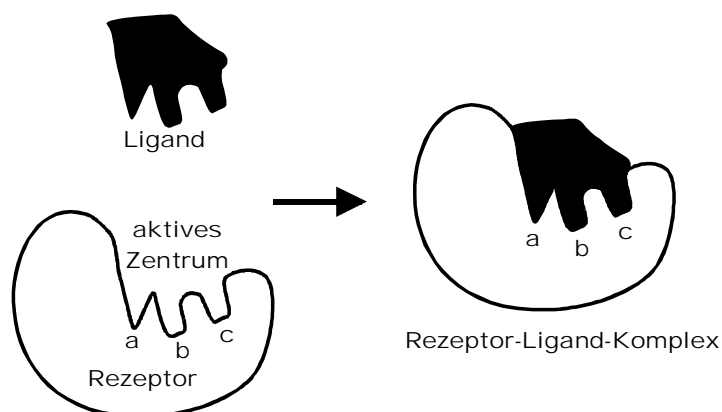


Abb. 2.1: Schlüssel-Schloß-Modell für die Rezeptor-Ligand-Wechselwirkung. Das aktive Zentrum des Rezeptors hat eine zum Liganden komplementäre Gestalt [7].

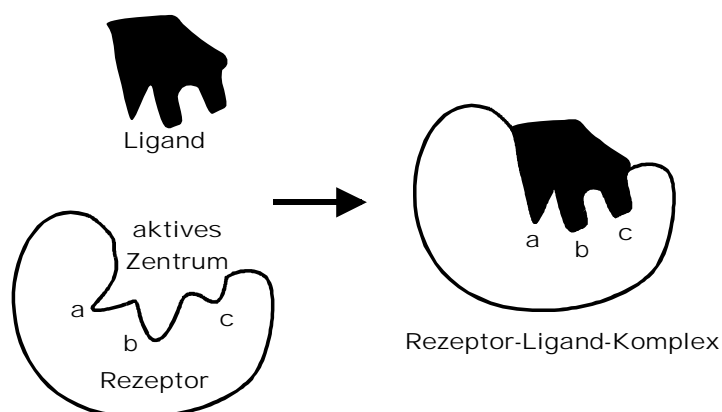


Abb. 2.2: *Induced Fit*-Modell für die Rezeptor-Ligand-Wechselwirkungen. Während der Entstehung des Komplexes bildet sich die Komplementarität des aktiven Zentrums des Rezeptors und des Liganden aus [7].

## 2.2 Docking-Algorithmen

Die mathematische Behandlung der molekularen Erkennung in biochemischen Komplexen spielt wie oben beschrieben eine Schlüsselrolle in Untersuchungen der katalytischen Wirkung von Enzymen und der Entwicklung neuer Wirkstoffe (*de novo*-Design). Daher wurde in der Literatur eine Vielzahl von Algorithmen für Untersuchungen auf dem Gebiet der Rezeptor-Ligand-Wechselwirkungen vorgestellt [9-13]. Viele dieser Algorithmen setzen voraus, daß die Bindungsstelle des Rezeptors, das aktive Zentrum, bekannt ist [14-18]. Die Liganden werden bei diesen Ansätzen dann möglichst optimal in dieses aktive Zentrum eingepaßt. Der wohl bekannteste und erste nicht auf Energieminimierung basierende *Docking*-Algorithmus überhaupt ist im Programm *DOCK* aus der Gruppe von Kuntz [19] implementiert. Dieses Programm erzeugt durch einander überlappende Kugeln ein negatives Bild des Rezeptors. Dieses Negativbild wird mit vier bis fünf Atomen des Liganden kombiniert, um eine eindeutige Orientierung des Liganden in der *Docking*-Region des Rezeptors zu erhalten. Weiterentwicklungen des Originalprogramms ermöglichen eine Energieminimierung der ermittelten Strukturen mit unterschiedlichen Minimierungsalgorithmen (genetische Algorithmen [20] oder *Downhill Simplex*-Minimierer [21]). In vielen Enzym-Ligand-Komplexen liegt das aktive Zentrum in einer großen konkaven Region. Ist das aktive Zentrum des Rezeptors nicht bekannt, kann deshalb die Vorhersage der Komplexstruktur versucht werden, indem der Algorithmus auf die größte oder mehrere große Kavitäten angewendet wird [22].

In Rahmen der vorliegenden Arbeit wird ein anderes Ziel verfolgt: Es wird angestrebt, die biologisch aktive Komplexstruktur ohne Vorkenntnisse über die Bindungsregion zu

bestimmen. In der folgenden Literaturübersicht wird deshalb auf Algorithmen eingegangen, die einen ähnlichen Ansatz verfolgen.

Die erste Gruppe von Algorithmen verwendet die sterische Komplementarität der molekularen Oberfläche (siehe Kapitel 3) als Kriterium für mögliche Komplexstrukturen. Connolly [23] berechnete für jeden Punkt dieser Oberfläche einen die Konkavität/Konvexität der Oberfläche charakterisierenden Wert. Dieser als *Solid Angle* bezeichnete Wert gibt den Anteil einer Kugel mit definiertem Radius an, der sich innerhalb des Moleküls befindet. Die lokalen Extrema des *Solid Angle* entsprechen dann entweder einem Knopf oder einem Loch der Topographie der molekularen Oberfläche. Durch die Kombination von vier Loch-Knopf-Paaren gegenüberliegender Oberflächen kann eine mögliche Komplexstruktur ermittelt werden. Hendrix und Kuntz [24] griffen diese Idee auf, indem sie Oberflächenbereiche mit komplementären Werten des *Solid Angle* bestimmten. Zur Ermittlung der *Docking*-Orientierungen wurden diese Oberflächenbereiche dann in einer adaptierten Version des Programmes *DOCK* verwendet. Lin et al. [25] entwickelten eine Einteilung der molekularen Oberfläche in konkave und konvexe Bereiche bzw. Sattelbereiche, die durch einen repräsentativen Punkt und einen mittleren Normalenvektor charakterisiert werden. Über einen als *Geometric Hashing* bezeichneten Algorithmus [26,27] werden Komplexstrukturen bestimmt, bei denen gleichzeitig mehrere konkave Bereiche des Rezeptors mit konvexen Bereichen des Liganden kombiniert werden können [28]. Die Flexibilität des Liganden bzw. des Rezeptors werden durch die Rotation größerer Molekülteile in einem erweiterten Algorithmus berücksichtigt [29,30]. Zwei weitere sehr ähnliche Ansätze benutzen direkt die Eigenschaften der molekularen Oberfläche. Bacon und Moulton [31] verwendeten die Lage der Oberflächenpunkte, Wallqvist und Covell [32] ermittelten aus drei geometrisch kompatiblen Paaren von Oberflächennormalen eine mögliche Komplexstruktur.

Zwei weitere Algorithmen versuchen, das *Docking*-Problem durch Unterteilung in zweidimensionale Teilprobleme zu lösen. Von Helmer-Citterich et al. [33,34] wurde ein Algorithmus vorgestellt, in dem die Oberfläche durch eine Menge von Polygonen und die Topographie dieser Polygone durch eine zweidimensionale Matrix beschrieben werden. Die Ermittlung von komplementären Oberflächenregionen erfolgt durch den Vergleich von Teilmatrizes. Badel et al. [35] bestimmten mögliche Komplexstrukturen durch Vergleich des Winkelprofils der zweidimensionalen Polygone der beiden Moleküle.

Ein in letzter Zeit wieder stark ins Interesse gerückter Ansatz ist der Fourier-Korrelations-Algorithmus von Katchalski-Katzir et al. [36]. Das Molekül wird dabei durch ein Gitter repräsentiert. Jedem Gitterpunkt wird der korrespondierende Wert für die Oberfläche, das Molekülinnere und den freien Raum zugeordnet. Die Komplementarität der Oberfläche wird bestimmt, indem die Fourier-Korrelation der Gitter beider Moleküle in jeder möglichen

Rotation der Moleküle zueinander berechnet wird. Die Effizienz der Methode ergibt sich durch die Verwendung der schnellen Fourier-Transformation. Sternberg et al. [37] entwickelten eine Energiefunktion für die Fourier-Korrelations-Methode, die neben der reinen sterischen Komplementarität die elektrostatische Komplementarität der Moleküle berücksichtigt. Dadurch sind vor allem bei *Docking*-Simulationen, die von den Strukturen der ungebundenen Komplexpartner ausgehen, bessere Ergebnisse zu erzielen. In einer weiteren Alternative wird der Abstand zwischen Wasserstoffbrückenbindungen für die Vorauswahl möglicher *Docking*-Strukturen verwendet [38]. Dabei werden jeder potentielle Wasserstoffbrückendonator mit jedem Wasserstoffbrückenakzeptor des anderen Moleküls kombiniert und die beiden Moleküle so zueinander positioniert, daß sich mindestens zwei Wasserstoffbrücken ausbilden können. Die so erhaltenen Strukturen werden mit der Funktion von Katchalski-Katzir et al. [36] auf sterische Überlappungen überprüft und energieminiert.

In allen bisher beschriebenen Verfahren wird zuerst durch einen empirischen Ansatz die mögliche Struktur der Komplexe bestimmt, und diese wird dann durch unterschiedliche Funktionen energetisch bewertet. Dagegen versuchen die Energieminimierungsverfahren durch die globale Optimierung einer Energiefunktion die Komplexstruktur zu ermitteln. Dabei unterscheiden sich die einzelnen Verfahren durch die verwendete Energiefunktion und die globale Optimierungsmethode. Neben der Monte-Carlo-Methode [39-42] und *Simulated Annealing* [43-45] werden in letzter Zeit immer häufiger genetische Algorithmen [46-50] verwendet. Der Vorteil der Verfahren zur Energieminimierung besteht in der Möglichkeit, die molekulare Flexibilität explizit zu berücksichtigen. Der größte Nachteil ist die sehr lange Rechenzeit.

### 2.3 *Fuzzy Set Theory* und *Fuzzy Logic*

Der Begriff *Fuzzy Set Theory* wurde 1965 von Lofty A. Zadeh [5] als Erweiterung der gewöhnlichen Mengenlehre eingeführt. Die *Fuzzy Logic* soll die unscharfen Begriffe menschlichen Denkens der Mathematik sowie der Technik erschließen [51]. Erst in den letzten Jahren wurde das Potential dieser Methode für chemische Fragestellungen erkannt [52,53]. Ihr Haupteinsatzgebiet liegt bei der *Fuzzy*-Regelung von großtechnischen Anlagen wie dem Polymerisationsprozeß [52,54] und dem Vergleich gemessener Spektren mit den in Datenbanken gespeicherten Spektren [55-58].

In der theoretischen Chemie wird die *Fuzzy Logic* zur Beschreibung der Chiralität [53,59], der Symmetrie [60] sowie der Form und Größe von Molekülen [53,61,62] verwendet. Daneben wurden zwei Ansätze zur Beschreibung des *Docking*-Problems mit Hilfe von *Fuzzy Logic*-Ansätzen entwickelt [59,63,64]. Der erste beschreibt die Korrelation der über *Fuzzy Logic*

quantifizierten Größe der Chiralität eines Substratmoleküls mit der Bindungskonstanten im aktiven Zentrum [59]. Die *Fuzzifizierung* der molekularen Oberfläche und eine daraus ermöglichte Ermittlung von Komplexstrukturen wird im zweiten Ansatz vorgestellt [63,64], der als Grundlage für den in dieser Arbeit vorgestellten Algorithmus dient. Die benötigten Grundlagen der *Fuzzy Set Theory* werden in diesem Kapitel kurz referiert. Ausführliche Darstellungen finden sich in der Literatur [3,4].

### 2.3.1 Definition von *Fuzzy*-Mengen

In der klassischen Mengenlehre gehört jeder Wert der Grundmenge entweder zur Menge oder zur entsprechenden Komplementärmenge. Diese Einteilung kann durch eine charakteristische Funktion, im weiteren Zugehörigkeitsfunktion (*Membership Function*) genannt, beschrieben werden. Bei Elementen, die zur Menge  $M$  gehören, nimmt diese Funktion den Wert 1 an, bei allen anderen Elementen hat sie den Wert 0. Im weiteren soll zur Unterscheidung von *Fuzzy*-Mengen dieser klassische Mengenbegriff als scharfe Menge (*Crisp Set*) bezeichnet werden.

$$\mu_M(x) = \begin{cases} 1 & \text{für } x \in M \\ 0 & \text{sonst} \end{cases} \quad (2.1)$$

mit

$x$ :

$M$ : scharfe Menge

$\mu_M(x)$ : Wert der Zugehörigkeitsfunktion von  $x$  zur Menge  $M$

In *Fuzzy*-Mengen (unscharfe Menge, *Fuzzy Set*) kann die über die Grundmenge definierte Zugehörigkeitsfunktion beliebige positive Werte annehmen. Dabei wird der Wertebereich meistens auf das Intervall  $[0;1]$  beschränkt und kann als Maß der Zugehörigkeit zur betreffenden Menge interpretiert werden. Ein Wert von 1 entspricht einer absoluten Zugehörigkeit zur Menge, ein Wert von 0 keiner Zugehörigkeit. Die *Fuzzy*-Menge  $\tilde{A}$  wird durch alle geordneten Paare der Elemente der Grundmenge und den dazugehörenden Werten der Zugehörigkeitsfunktion beschrieben:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid x \in X\} \quad (2.2)$$

mit

$X$ : Grundmenge

$x$ : Elemente der Grundmenge

$\tilde{A}$ : *Fuzzy*-Menge

$\mu_{\tilde{A}}(x)$ : Wert der Zugehörigkeitsfunktion von  $x$  zur *Fuzzy*-Menge  $\tilde{A}$

Alle Elemente der Grundmenge, deren Zugehörigkeit zur *Fuzzy*-Menge größer oder gleich einem bestimmten Wert  $\alpha$  ist, werden als  $\alpha$ -Niveaumenge ( $\alpha$ -Cut,  $\alpha$ -Level Set) definiert:

$$A^\alpha = \{x \in X \mid \mu_{\tilde{A}}(x) \geq \alpha\} \quad (2.3)$$

mit

$A^\alpha$ :  $\alpha$ -Niveaumenge  $\alpha \in [0;1]$

Die scharfe Menge  $A_{\text{strong}}^\alpha = \{x \in X \mid \mu_{\tilde{A}}(x) > \alpha\}$  wird als strenge  $\alpha$ -Niveaumenge (*Strong  $\alpha$ -Cut*) bezeichnet. Der Spezialfall einer strengen  $\alpha$ -Niveaumenge mit  $\alpha = 0$  heißt Träger (Stütze, *Support*) der *Fuzzy*-Menge.

### 2.3.2 Linguistische Variable

Eine besondere Stärke der *Fuzzy Set Theory* besteht in der Einbeziehung von Begriffen der menschlichen Sprache in die Mathematik. Dies erfolgt über die Einführung von *linguistischen Variablen*. Diese werden als Gruppen von *Fuzzy*-Mengen mit einander überlappenden Zugehörigkeitsfunktionen der gleichen Basisvariablen definiert. Die unscharfen Mengen einer linguistischen Variablen stehen als beschreibende Eigenschaften verschiedener Teilaspekte des übergeordneten Begriffs der linguistischen Variablen in einem begrifflichen Zusammenhang zueinander und lassen sich als *Klassen* der linguistischen Variablen verstehen. Eine linguistische Variable  $\mathcal{L}$  kann somit mit  $n$  ( $n \geq 1$ ) sie klassifizierenden *Fuzzy*-Mengen  $\tilde{A}_1, \dots, \tilde{A}_n$  definiert werden:

$$\mathcal{L} = \{\tilde{A}_1, \dots, \tilde{A}_n\} \quad (2.4)$$

oder 
$$\mathcal{L} = \{(x, \mu_{\tilde{A}_1}(x)), \dots, (x, \mu_{\tilde{A}_n}(x))\} \quad (2.5)$$

Die einzelnen Klassen der linguistischen Variablen werden zumeist durch *Fuzzy*-Zahlen oder *Fuzzy*-Intervalle gebildet. Diese unscharfen Beschreibungen entsprechen einer verbalen Ausdrucksweise wie "ungefähr gleich  $x$ " bzw. "ungefähr zwischen  $x$  und  $y$ ". Eine *Fuzzy*-Zahl



besteht aus einem ansteigenden Ast, bis die Zugehörigkeitsfunktion eines Elements der Grundmenge den Wert 1 annimmt. Anschließend fällt die Zugehörigkeitsfunktion wieder auf 0 ab (Abb. 2.3). Bei einem *Fuzzy*-Intervall liegen zwischen dem aufsteigenden und dem absteigenden Ast ein geschlossenes Intervall der Grundmenge mit einem Wert der Zugehörigkeitsfunktion von 1 (Abb. 2.3).

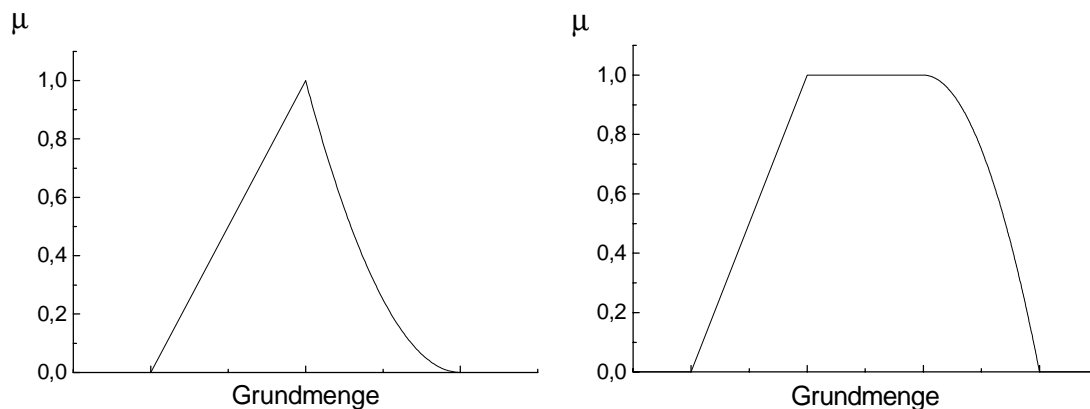


Abb. 2.3: Schematische Darstellung einer *Fuzzy*-Zahl und eines *Fuzzy*-Intervalls

### 2.3.3 Verknüpfungsoperatoren von *Fuzzy*-Mengen

Der Entscheidungsfindungsprozeß mit Hilfe der *Fuzzy Logic* gliedert sich in drei Teile:

- *Fuzzifizierung* reeller Werte
- Verknüpfung (*Inferenz*) verschiedener linguistischer Variablen mit diversen logischen Operatoren
- *Defuzzifizierung* zu einem reellen Wert

Der erste Schritt, die *Fuzzifizierung*, entspricht einer Zuordnung der scharfen Meßwerte und Eigenschaften zu *Fuzzy*-Mengen bzw. linguistischen Variablen. Anschließend erfolgt die Auswertung durch logische Verknüpfungsoperationen (*Inferenz*) dieser *Fuzzy*-Daten. Um eine eindeutige Entscheidung zu fällen, muß für die meisten Fragestellungen das wiederum als *Fuzzy*-Wert vorliegende Ergebnis in einen scharfen Wert zurückübersetzt werden. Dieser letzte Teilschritt wird als *Defuzzifizierung* bezeichnet.

Bei der Entscheidungsfindung sind vor allem die logischen Junktoren von entscheidender Bedeutung. Für diese Verknüpfungen wurde eine Vielzahl von Operationen beschrieben [3,4]. Dabei wurde darauf geachtet, daß die klassische Mengenlehre bzw. Boole'sche Logik als

Spezialfall enthalten ist. Hier soll nur auf die von Zadeh [5] vorgestellten Operationen eingegangen werden, die die Junktoren "und", "oder" und "nicht" als Schnitt-, Vereinigungs- bzw. Komplementärmenge wie folgt definieren:

- *Fuzzy-Schnittmenge*  $\tilde{C} = \tilde{A} \cap \tilde{B} \rightarrow \text{Minimum-Operator}$   
mit  $\mu_{\tilde{C}}(x) = \mu_{\tilde{A} \cap \tilde{B}}(x) = \min(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x))$  (2.6)

- *Fuzzy-Vereinigungsmenge*  $\tilde{C} = \tilde{A} \cup \tilde{B} \rightarrow \text{Maximum-Operator}$   
mit  $\mu_{\tilde{C}}(x) = \mu_{\tilde{A} \cup \tilde{B}}(x) = \max(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x))$  (2.7)

- *Fuzzy-Komplementärmenge*  $\tilde{C} = \neg \tilde{A}$   
mit  $\mu_{\tilde{C}}(x) = \mu_{\neg \tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x)$  (2.8)

Von Heiden et al. [63,65] wurde eine Funktion vorgestellt, die den Vergleich von linguistischen Variablen ermöglicht. Diese Unähnlichkeitsfunktion  $D_{LV}(\mathcal{A}, \mathcal{B})$  (*Dissimilarity Function*) beschreibt die Ähnlichkeit bzw. Unähnlichkeit zweier linguistischen Variablen des gleichen Typs. Dabei können die verschiedenen Klassen der linguistischen Variablen über die Wichtungsfaktoren  $w_i$  unterschiedlich stark berücksichtigt werden.

$$D_{LV}(\mathcal{A}, \mathcal{B}) = \frac{\sum_{i=1}^n w_i \cdot |\mu_{\tilde{A}_i}(x) - \mu_{\tilde{B}_i}(x)|}{\sum_{i=1}^n w_i \cdot (\mu_{\tilde{A}_i}(x) + \mu_{\tilde{B}_i}(x))} \quad (2.9)$$

mit

$\mathcal{A}, \mathcal{B}$ : Linguistische Variablen identischen Typs:

$$\mathcal{A} = \{(x, \mu_{\tilde{A}_1}(x)), \dots, (x, \mu_{\tilde{A}_n}(x))\}$$

$$\mathcal{B} = \{(x, \mu_{\tilde{B}_1}(x)), \dots, (x, \mu_{\tilde{B}_n}(x))\}$$

$w_i$ : Wichtungsfaktor der Klasse  $i$ ,  $0 \leq w_i \leq 1$

$N$ : Anzahl der Klassen in  $\mathcal{A}, \mathcal{B}$

Es ergibt sich für die Unähnlichkeitsfunktion ein Wertebereich von 0 bis 1, wobei ein Wert von 0 vollkommene Identität und ein Wert von 1 größtmögliche Unähnlichkeit kennzeichnet. Als Ähnlichkeit zweier linguistischer Variablen  $S_{LV}(\mathcal{A}, \mathcal{B})$  wird hier das Komplement der Unähnlichkeitsfunktion definiert.

$$S_{LV}(\mathcal{A}, \mathcal{B}) = 1 - D_{LV}(\mathcal{A}, \mathcal{B}) \quad (2.10)$$

## 2.4 Genetische Algorithmen

In Anlehnung an die Natur, in der die am besten angepaßten Individuen einer Population die größten Chancen zum Überleben besitzen und ihre Eigenschaften an Nachkommen vererben können, propagieren genetische Algorithmen (GA) Lösungen von Generation zu Generation und kombinieren sie, um weitere Verbesserungen zu erreichen. So vereinen genetische Algorithmen die Prinzipien der natürlichen Auslese und der biologischen Evolution zu effektiven Optimierungsalgorithmen [66,67].

Der Startpunkt eines genetischen Algorithmus ist eine Population von Individuen, wobei jedes Individuum eine mögliche Lösung des untersuchten Problems darstellt. Im Fall der molekularen Erkennung codiert ein solches Individuum die momentane Anordnung der beiden Komplexpartner. Als Parameter für den Optimierungsalgorithmus werden dabei die Translation und Rotation der beiden Moleküle relativ zueinander sowie die Konformationen der Moleküle verwendet. Zur Codierung werden üblicherweise binäre Vektoren (im Englischen als *Bit Strings* bezeichnet) verwendet. Jeder Eingabeparameter wird durch eine bestimmte Anzahl von *Bits* repräsentiert, wobei jeder durch diesen binären Vektor kodierte Wert einem bestimmten Wert der meist als reelle Zahlen vorliegenden Eingabeparameter entspricht. Andererseits können auch direkt die reellen Zahlen zu einem Vektor zusammengefaßt werden. Der gesamte Vektor wird auch in Anlehnung an die Natur als Chromosom, die Codierung jedes einzelnen Eingabeparameters als Gen bezeichnet.

Nach der Initialisierung der Chromosomen der gesamten Population beginnt der iterative Prozeß des genetischen Algorithmus. Dazu werden auf die einzelnen Chromosomen eine Reihe von Mechanismen angewendet, wodurch die Elternpopulation durch eine neue Population, die Nachkommen, ersetzt wird. Diese Mechanismen werden als Selektion, Crossover und Mutation bezeichnet.

### 2.4.1 Selektion

Als erster Schritt wird für jedes Individuum bestimmt, wie gut die Parameter dieses Chromosoms die gestellte Optimierungsaufgabe lösen. Dies erfolgt über eine zu dem speziellen Problem gehörende Bewertungsfunktion und wird als *Fitness* bezeichnet. Daran anschließend werden einzelne Chromosomen durch ein Zufallsprinzip ausgewählt. Dabei bestimmt die *Fitness* eines bestimmten Chromosoms im Verhältnis zu allen anderen die Wahrscheinlichkeit, mit der dieses Chromosom bei der Selektion zur Auswahl kommt. Chromosomen, die einer besseren Lösung entsprechen, werden öfter ausgewählt als schlechtere. Dadurch soll gewährleistet werden, daß die Generation der Nachkommen

insgesamt eine bessere *Fitness* als die Elterngeneration besitzt, wodurch das biologische Prinzip “*Survival of the Fittest*” nachgeahmt werden soll. Die so ausgewählten Chromosomen werden dann weiteren Operationen unterzogen.

### 2.4.2 Crossover

Beim Crossover werden Teile der Chromosomen paarweise zwischen Individuen ausgetauscht. Dadurch wechseln Teillösungen des Problems von einem Individuum zu einem anderen. Dies entspricht dem biologischen Vorgang der Kreuzung. Von den vielen verschiedenartigen Crossover-Operatoren werden das Ein-Punkt- bzw. das Zwei-Punkt-Crossover am häufigsten verwendet.

Das Ein-Punkt-Crossover trennt zwei Chromosomen an der gleichen, zufällig ausgewählten Stelle. Die beiden hinteren Teilstücke werden getauscht, so daß die Nachkommen jeweils aus dem vorderen Teil des einen und dem hinteren Teil des anderen Chromosoms bestehen. Beim Zwei-Punkt-Crossover erfolgt die Trennung an zwei zufällig ausgewählten Stellen, und die Mittelstücke der Chromosomen werden ersetzt.

### 2.4.3 Mutation

Mutationen werden verwendet, um in den genetischen Prozeß Merkmale einzubringen, die bis zu dem jeweiligen Zeitpunkt kein Mitglied der aktuellen Population besitzt. Dadurch sorgt die Mutation für Divergenz und Inhomogenität innerhalb der Population, so daß möglichst der gesamte Konfigurationsraum abgesucht und eine frühzeitige Konvergenz verhindert wird. Aus einem zu mutierenden Chromosom wird per Zufall ein Vektorelement ausgewählt. Bei einer einfachen binären Repräsentation der Eingabeparameter erfolgt die Mutation durch Invertierung der ausgewählten *Bits*. Werden reelle Zahlen zur Codierung verwendet, kann eine Mutation entweder durch Addition bzw. durch Subtraktion eines konstanten Wertes oder eines zufällig gewählten Wertes aus einem bestimmten Intervall erfolgen.

### 2.4.4 Weitere Operatoren genetischer Algorithmen

Für eine schnellere und gesicherte Konvergenz des genetischen Algorithmus wurden in dieser Arbeit neben den oben beschriebenen Grundoperationen zwei weitere verwendet, die kurz vorgestellt werden sollen:

Für die lokale Optimierung eines Problems sind genetische Algorithmen wenig geeignet. Wurde durch den Algorithmus erst einmal eine Lösung in der Nähe eines Minimums identifiziert, kann durch eine lokale Optimierungsmethode das Auffinden des lokalen Minimums sehr stark beschleunigt werden. Dadurch können die Vorteile der globalen Optimierung des genetischen Algorithmus und der lokalen Optimierung kombiniert werden. Im einfachsten Ansatz kann das durch den genetischen Algorithmus ermittelte Minimum durch einen lokalen Optimierer, z.B. die *Downhill Simplex*-Methode [68], weiter verbessert werden. Andererseits kann der lokale Optimierer auch im genetischen Algorithmus integriert werden. Dazu können in jeder Generation alle oder nur ein Teil der durch die Individuen repräsentierten Lösungen weiter optimiert werden, wodurch eine schnellere Konvergenz erzielt wird [69].

Mit der zweiten Operation soll garantiert werden, daß das beste Individuum der Eltern-Population auch in der nachfolgenden Generation enthalten ist [69]. Dazu wird eine Kopie des besten Chromosoms direkt als Nachkomme übernommen. Dieses als *Elitist Generational Replacement Genetic Algorithm* bezeichnete Verfahren [69] garantiert, daß ein einmal gefundenes globales Minimum in späteren Generationen nicht mehr verloren geht.

## Kapitel 3

### Topographische und physikochemische Eigenschaften molekularer Oberflächen

Der in dieser Arbeit entwickelte *Docking*-Algorithmus beruht auf dem Modell einer molekularen Oberfläche und auf der Zuordnung topographischer und physikochemischer Eigenschaften zu jedem Punkt dieser Oberfläche. Deshalb werden hier das verwendete Modell einer molekularen Oberfläche und anschließend die verwendeten topographischen und physikochemischen Eigenschaften sowie ihre Berechnung am Ort der Oberfläche bzw. ihre Projektion auf die Oberfläche beschrieben.

#### 3.1 Molekulare Oberflächen

Moleküle lassen sich durch eine Elektronendichteverteilung, die mit zunehmendem Abstand gegen null strebt, beschreiben [70]. Eine molekulare Oberfläche im makroskopischen Sinn hat sich jedoch als Modellvorstellung bei der Untersuchung der Wechselwirkungen zweier Komplexpartner als sehr nützlich erwiesen, da sie einen Eindruck über Größe, Ausdehnung und topographische Eigenschaften der betrachteten Moleküle vermitteln kann. Für die Definition einer molekularen Oberfläche wurden in der Literatur mehrere verschiedene Ansätze vorgestellt [70-76], wobei die van-der-Waals-Oberfläche [71], die *Cavity Surface* [77] sowie die *Solvent Accessible Surface* [73,74] zu den wichtigsten zählen. Im Folgenden wird auf die in dieser Arbeit verwendeten *Solvent Accessible Surface* näher eingegangen.

Die Definition der *Solvent Accessible Surface*, auch oft als Connolly-Oberfläche bezeichnet, beruht auf dem van-der-Waals-Atommodell. Darin werden die einzelnen Atome als Kugeln mit vordefinierten Radien, die sich aus dem Mindestabstand zweier nichtbindender Atome ergeben, beschrieben. In dieser Arbeit werden die von Bondi [78] ermittelten van-der-Waals-Radien verwendet. Über dieses Modell wird ein kugelförmiges Probeteilchen gerollt (Abb. 3.1). Die Connolly-Oberfläche ergibt sich aus der Kontaktfläche zwischen der Probekugel und den van-der-Waals-Kugeln (*Contact Surface*) und einer interpolierten Oberfläche (*Reentrant Surface*). Die *Reentrant Surface* wird durch die dem Molekül zugewandten Oberflächenbereiche der Probekugel gebildet, wenn diese gleichzeitig mehrere Atome berührt. Die *Contact Surface* entspricht somit den konvexen Oberflächenbereichen und die *Reentrant Surface* den konkaven Bereichen, wenn die Probekugel mit mehr als zwei Atomen wechselwirkt. Besteht Kontakt der Probekugel mit zwei Atomen, entspricht dies Sattelbereichen der Oberfläche. Für die Probekugel findet in der Regel der Radius eines als

kugelförmig angenommenen Wassermoleküls von 1.4 Å Verwendung. Damit repräsentiert das Volumen innerhalb der *Solvent Accessible Surface* den Raum, der vom Lösungsmittel und damit auch von möglichen Komplexpartnern nicht erreicht werden kann.

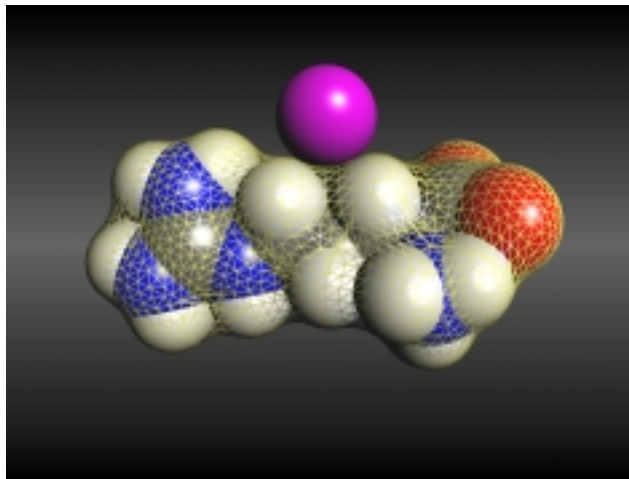


Abb. 3.1: Darstellung der Generierung einer molekularen Oberfläche [79]

Die violette Kugel repräsentiert das Probeteilchen und das gelbe Dreiecksnetz die triangulierte Connolly-Oberfläche

Die Berechnung der Oberfläche als Punktoberfläche kann mit dem Standard-Computerprogramm von Connolly (*Molecular Surface MS* [QCPE 429]) erfolgen [74,75]. Zum Zweck der graphischen Darstellung kann das so erzeugte Punktraster nachträglich durch Triangulation in ein zusammenhängendes Dreiecksnetz überführt werden [80]. Eine weitere Methode zur Generierung von Connolly-Oberflächen ist das von Sanner et al. [81] entwickelte Programm *MSMS*, das mit Hilfe eines analytischen Algorithmus direkt eine triangulierte Oberfläche erzeugt.

### 3.2 Topographische Analyse der molekularen Oberfläche

Biochemische Komplexe entstehen durch die Ausbildung attraktiver intermolekularer Wechselwirkungen bei der Annäherung der beteiligten Moleküle. Die dazu nötige sterische Komplementarität wird in dieser Arbeit durch topographische Eigenschaften der molekularen Oberfläche der einzelnen Moleküle beschrieben. So kann der Vergleich der Komplementarität unabhängig von der relativen Lage der Moleküle zueinander erfolgen.

### 3.2.1 Kanonische und globale Krümmungen

Bei der Untersuchung der topographischen Eigenschaften einer Oberfläche ist zunächst die effektive Auflösung (Größenordnung) der Analyse zu definieren [82,83]. Eine Oberfläche kann bei atomarer Auflösung viele topographische Charakteristika zeigen, die bei Verringerung der Auflösung verschwinden. Gerade die Vernachlässigung dieser nur bei atomarer Auflösung auftretenden Merkmale kann aber zur Charakterisierung der Eigenschaften größerer Oberflächenbereiche verwendet werden.

Die lokalen, topographischen Eigenschaften der molekularen Oberfläche können durch die beiden *kanonischen Krümmungen* quantifiziert werden [84,85]. Die kanonischen Krümmungen sind als Eigenwerte der lokalen Hesse-Matrix definiert und beschreiben die unmittelbare Umgebung eines Oberflächenpunktes. Somit beschreiben die kanonischen Krümmungen die Topographie des Moleküls in atomarer Größenordnung. Konvexe Krümmungen sind negativ, konkave positiv definiert. Die Topographie der Oberflächenpunkte wird in konvexen Regionen durch zwei negative, die in konkaven Regionen durch zwei positive Krümmungen beschrieben. Sattelpunkte besitzen eine negative und eine positive kanonische Krümmung.

Für die molekulare Erkennung ist jedoch die Komplementarität größerer, molekularer Fragmente verantwortlich, so daß die globale Beschreibung größerer Oberflächenbereiche die Bindungsregionen besser charakterisieren sollte. Von Zachmann [86,87] wurde die Definition der kanonischen Krümmungen deswegen auf Oberflächenbereiche vordefinierter Größe erweitert. Zur Berechnung dieser globalen Krümmungen werden zu jedem Oberflächenpunkt in einem frei wählbaren Abstand entlang der Oberfläche Grenzpunkte definiert und an diese Grenzpunkte mit Hilfe eines numerischen Verfahrens ein Paraboloid angenähert. Die Hauptkrümmungen des Paraboloids definieren die globalen Krümmungen an jedem Oberflächenpunkt [86,87]. Dabei wird die jeweils größere Hauptkrümmung eines Oberflächenpunktes als erste globale Krümmung (*global curvature 1*,  $k_1$ ) und die kleinere als zweite globale Krümmung (*global curvature 2*,  $k_2$ ) bezeichnet.

Da sich mit dieser Berechnungsmethode starke Unterschiede in den Werten benachbarter Oberflächenpunkte ergeben (Abb. 3.2), was auf eine starke Abhängigkeit von der Lage und Definition der Grenzpunkte schließen läßt, wird in dieser Arbeit die Definition der globalen Krümmungen modifiziert. So werden nicht nur die Punkte im vordefinierten Abstand als Referenzpunkte für die Bestimmung des Paraboloids herangezogen, sondern sämtliche Punkte bis zu diesem vordefinierten Abstand. Dazu wird zuerst eine Nachbarschaftsliste jedes Oberflächenpunktes erstellt. Als nächste Nachbarn werden die Punkte bezeichnet, die durch einen Seite eines Oberflächendreiecks mit dem Zentralpunkt verbunden sind.



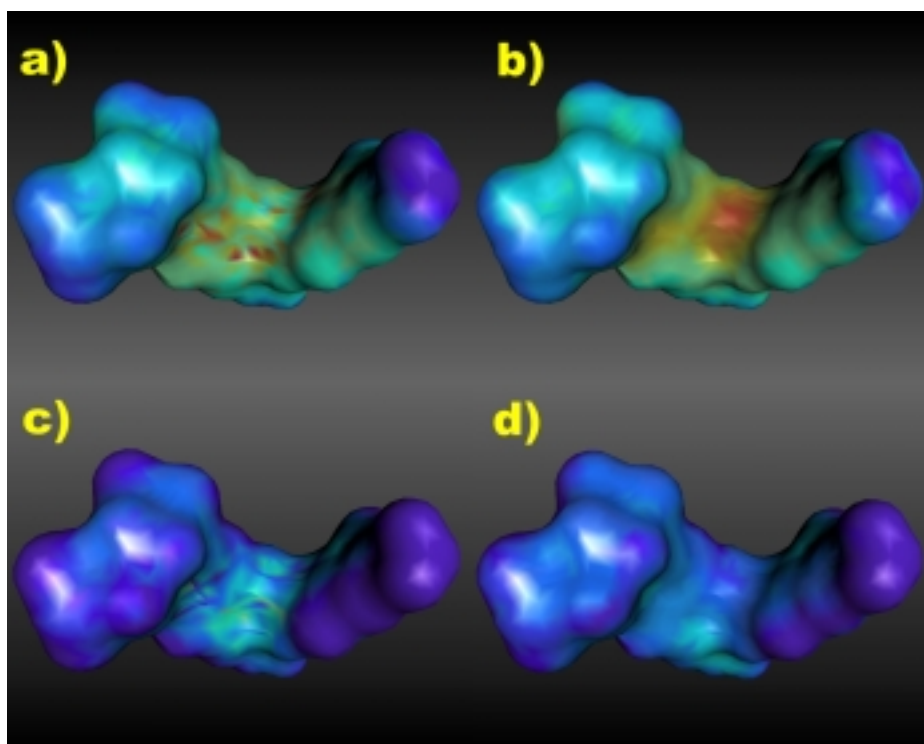


Abb. 3.2: Vergleich der globalen Krümmungen von Methotrexat. Die Werte der ersten globalen Krümmungen sind im oberen Teil, die der zweiten globalen Krümmung im unteren Teil farbcodiert auf der Oberfläche wiedergegeben. Dabei entsprechen rote Bereiche positiven Krümmungen. Negative Krümmungen sind blau eingefärbt.

a und c) Globale Krümmung 1 bzw. 2 nach der Definition von Zachmann [86,87]  
b und d) Globale Krümmung 1 bzw. 2 nach der modifizierten Definition

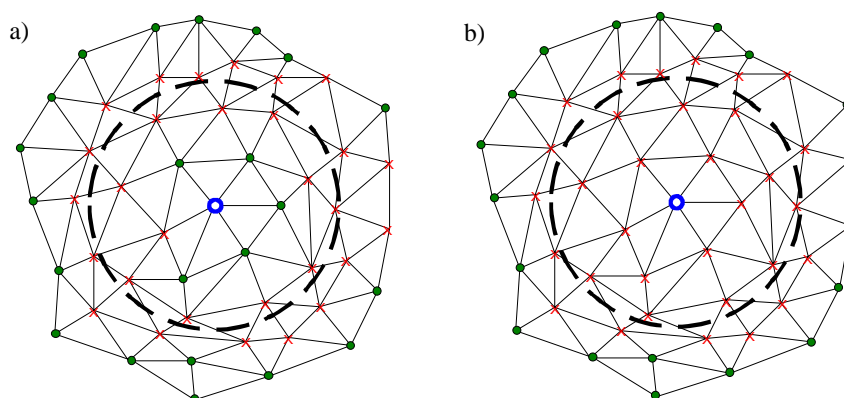


Abb. 3.3: Definition der Referenzpunkte für die Bestimmung des Paraboloids  
a) Definition nach Zachmann [86,87]

b) modifizierte Definition

- Oberflächenpunkt
- Zentralpunkt
- × Referenzpunkt
- vordefinierter Abstand

Ausgehend vom Zentralpunkt werden nun alle Nachbarpunkte rekursiv in die Liste der Referenzpunkte aufgenommen, bis der kürzeste Abstand entlang der Dreiecksseiten den vordefinierten Abstand zum Zentralpunkt überschreitet (Abb. 3.3). Dadurch wird ein kontinuierlicher Verlauf der globalen Krümmungen gewährleistet (Abb.3.2), da sich für benachbarte Zentralpunkte die Liste der Referenzpunkte und dadurch auch die durch diese Punkte bestimmten Paraboloiden nur leicht unterscheiden.

### 3.2.2 Beschreibung der Form und Neigung molekularer Oberflächen

Für die Beschreibung der Topographie von Oberflächen werden in der Differentialgeometrie mehrere Ansätze zur Kombination der beiden Krümmungen zu charakteristischen Zahlen beschrieben [88]. So ist die mittlere Krümmung als arithmetisches Mittel und die Gaußsche Krümmung als Produkt der beiden Krümmungen definiert (Gleichung 3.1 bzw. 3.2).

$$\text{mittlere Krümmung} \quad H = \frac{1}{2} \cdot (k_1 + k_2) \quad (3.1)$$

$$\text{Gaußsche Krümmung} \quad K = k_1 \cdot k_2 \quad (3.2)$$

Der hier vorgestellte *Docking*-Algorithmus (vor allem die Einteilung der Oberfläche in diskrete Bereiche) benötigt eine Zuordnung der topographischen Eigenschaften zu linguistischen Variablen (siehe Kapitel 3.6). Deswegen werden die Krümmungen der molekularen Oberflächen zu zwei charakteristischen topographischen Indizes kombiniert, von denen einer die Form (Konkavität / Konvexität) und der andere die Neigung (Stärke der Krümmung) der Oberfläche quantifiziert. Diese können dann direkt in linguistische Variablen überführt werden. Durch den ersten charakteristischen Wert wird das Verhältnis der Krümmungen, also die Konvexität / Konkavität, unabhängig vom Betrag der Krümmungen beschrieben. Der andere repräsentiert einen Mittelwert der Absolutwerte der beiden globalen Krümmungen.

In der Literatur wurden fast gleichzeitig zwei unterschiedliche Definitionen zur Beschreibung der Form der Oberfläche ausgehend von den Krümmungen vorgestellt [65,82]. Der von Heiden et al. [63,65] entwickelte *Surface Topography Index (STI)* beschreibt den kontinuierlichen Übergang der Oberflächenform zwischen fünf Eckwerten. Diese Eckwerte entsprechen gleichmäßig konkaven (Loch,  $STI=0$ ), langgestreckt konkaven (Spalt,  $STI=1$ ), sattelförmigen (Sattel,  $STI=2$ ), langgestreckt konvexen (Grat,  $STI=3$ ) bzw. gleichmäßig konvexen Bereichen (Pfropf,  $STI=4$ ). Der  $STI$  kann an jedem Punkt der Oberfläche aus den beiden globalen Krümmungen wie folgt berechnet werden [63,65]:

$$STI = \frac{k_1 - k_2}{k_1} \quad \text{wenn } k_1 > 0 \text{ und } k_2 > 0 \text{ oder wenn } k_1 > 0, k_2 \leq 0$$

$$\text{und } |k_1| > |k_2|$$

$$STI = \frac{k_1 + 3 \cdot k_2}{k_2} \quad \text{wenn } k_1 > 0 \text{ und } k_2 \leq 0, |k_1| \leq |k_2| \text{ oder wenn}$$

$$k_1 \leq 0 \text{ und } k_2 < 0$$

Der Sonderfall einer absolut ebenen Oberflächenregion, wenn also beide globalen Krümmungen gleich Null sind, wird mit einem *STI*-Wert von -1 berücksichtigt.

$$STI = -1 \quad \text{wenn } k_1 = k_2 = 0$$

Der zweite Ansatz stammt von Duncan und Olson [82,85] und wurde für die kanonischen Krümmungen entwickelt. Der als *Shape Index* (*s*) bezeichnete Zahlenwert ist aber auch analog auf globale Krümmungen anwendbar und wird nach Gleichung 3.3 berechnet:

$$s = -\frac{2}{\pi} \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right) \quad (3.3)$$

Die Werte des *Shape Index* laufen von  $s = -1$  für konkave Bereiche mit zwei gleich großen Krümmungen ( $\arctan(\infty) = \pi/2$ ) über 0 für Sattellbereiche mit betragsmäßig gleichen Krümmungen ( $\arctan(0) = 0$ ) nach  $s = +1$  für konvexe Bereiche ( $\arctan(-\infty) = -\pi/2$ ).

Trotz der stark unterschiedlichen Definition der beiden Topographie-Indizes zeigen sich nur geringe Unterschiede bei der farbcodierten Darstellung der Werte auf der Oberfläche (Abb. 3.4). So können die *STI*-Werte 0, 2 und 4 direkt einem *Shape Index* von -1, 0 bzw. +1 zugeordnet werden. Die Eckwerte Spalt ( $STI=1$ ) und Grat ( $STI=3$ ) entsprechen einem Wert von  $s = -0.5$  bzw.  $s = 0.5$  des *Shape Index*. Da der *Shape Index* durch seine geschlossene analytische Form leichter handhabbar ist, wird im folgenden nur noch diese Definition verwendet.

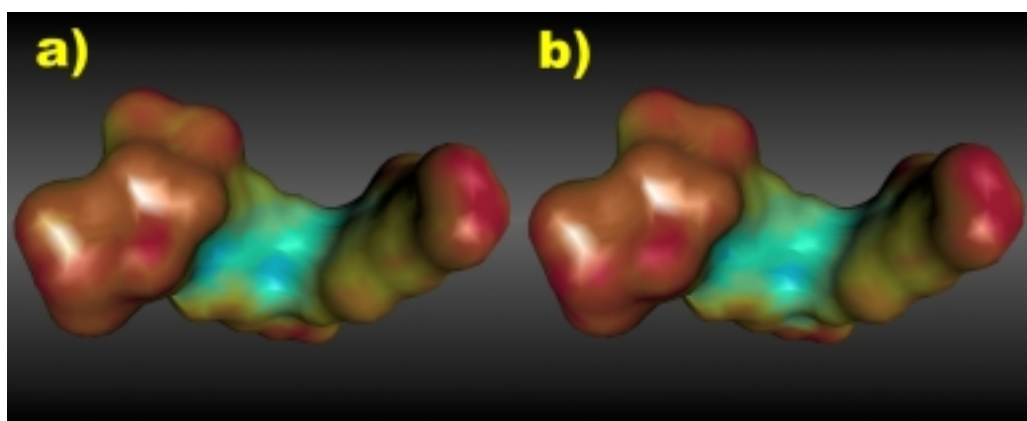


Abb. 3.4 Vergleich des *Surface Topography Index* (a) und des *Shape Index* (b). Die Werte sind jeweils farbcodiert auf der molekularen Oberfläche dargestellt. Rote Bereiche entsprechen einem *STI* von 4 bzw. einem *Shape Index* von 1 und blaue Bereiche einem *STI* von 0 bzw. einem *Shape Index* von -1.

Die Extremsituationen sehr große bzw. sehr kleine Krümmungen werden durch zwei Modifikationen behandelt.

Bei der Berechnung der Krümmungen können die Werte für einzelne Oberflächenpunkte sehr hohe Werte annehmen, die teilweise einer Krümmung im atomaren oder subatomaren Bereich entsprechen. Diese sehr hohen Werte haben aber einen sehr starken Einfluß auf die *Docking*-Ergebnisse, weswegen der für die Berechnung des *Shape Index* herangezogene Wertebereich auf das Intervall von  $-0.8 \text{ \AA}^{-1}$  bis  $0.8 \text{ \AA}^{-1}$  beschränkt wird. Alle Werte, die diesen Bereich unter- bzw. überschreiten, werden auf den kleinsten ( $-0.8 \text{ \AA}^{-1}$ ) bzw. größten Wert ( $0.8 \text{ \AA}^{-1}$ ) des Bereiches gesetzt.

Bei sehr kleinen Werten haben Ungenauigkeiten bei der Ermittlung der Krümmungen einen starken Einfluß auf den Wert des *Shape Index*. So kann es z.B. vorkommen, daß sich durch diese Fehler das Vorzeichen einer oder beider Krümmungen ändert, woraus auch eine Änderung des Vorzeichens des *Shape Index* resultiert. Liegen beide globalen Krümmungen betragsmäßig unter  $0.15 \text{ \AA}^{-1}$  wird der *Shape Index* über Gleichung 3.4 zu kleineren Werten hin korrigiert. Dadurch wird diesen Regionen der Wert für ein fast vollkommen ebenes Oberflächenstück zugeordnet:

$$s_{corr} = s \cdot \frac{|k_1| + |k_2|}{0.3 \text{ \AA}^{-1}} \quad (3.4)$$

mit

$s$ : *Shape Index* aus Gleichung 3.3

$s_{corr}$ : Korrigierter *Shape Index*

Für die Beschreibung der Neigung der Oberfläche wird in dieser Arbeit die von Duncan und Olson [82,85] eingeführte *Curvedness* ( $R$ ) verwendet, die als Wurzel aus dem quadratischen Mittelwert der beiden Krümmungen definiert ist:

$$R = \sqrt{\frac{1}{2} \cdot (k_{\max}^2 + k_{\min}^2)} \quad (3.5)$$

Bei der Berechnung der *Curvedness* wird wiederum nur das oben beschriebene Wertintervall der globalen Krümmungen berücksichtigt, so daß sich für die *Curvedness* ein Maximalwert von  $0.8 \text{ \AA}^{-1}$  ergibt.

Zwischen den bekannten Krümmungsmaßen Gaußsche Krümmung  $K$  und mittlere Krümmung  $H$  und dem hier verwendeten *Shape Index*  $s$  und der *Curvedness*  $R$  bestehen folgende Beziehungen:

$$s = -\frac{2}{\pi} \arctan\left(\frac{H}{\sqrt{H^2 - G}}\right) \quad (3.6)$$

$$R = \sqrt{H^2 - \frac{K}{2}} \quad (3.7)$$

### 3.3 Elektrostatisches Potential

Für eine stabile Komplexbildung müssen die beteiligten Moleküle neben der sterischen Komplementarität eine starke elektrostatische Komplementarität in den Bindungsregionen aufweisen. In dieser Arbeit wird die elektrostatische Komplementarität durch den Vergleich des elektrostatischen Potentials am Ort der molekularen Oberfläche bestimmt. Für kleine Moleküle läßt sich an jedem Punkt im Raum das elektrostatische Potential mit Hilfe von *ab-initio*- oder semiempirischen Methoden berechnen. Der für biochemische Makromoleküle benötigte sehr hohe Rechenaufwand ist für den hier vorgestellten Algorithmus nicht akzeptabel. Da biochemische Reaktionen im wäßrigen Medium ablaufen, muß außerdem das Lösungsmittel bei der Berechnung des elektrostatischen Potentials mit berücksichtigt werden. Dies kann mit Hilfe der Kontinuumstheorie erfolgen. Dieser Ansatz beruht darauf, daß die Atome durch punktförmige Partialladungen beschrieben werden, die entweder aus auf Fragmentierungsmethoden beruhenden Kraftfeldern entnommen oder mit Hilfe semiempirischer Methoden berechnet werden.

Die Kontinuumstheorie beschreibt das gelöste Molekül als Kavität niedriger dielektrischer Konstante (für Protein  $\epsilon = 2$  bis  $6$ ) in einem Medium mit hoher Dielektrizitätskonstante (für

Wasser  $\epsilon = 80$ ) [89,90]. Die Begrenzung der Kavität ist durch die *Solvent Accessible Surface* gegeben. Die Ladungsverteilung innerhalb des Moleküls wird durch die Partialladung an den Atommittelpunkten bestimmt, während die Solventphase durch eine konstante Ionenstärke approximiert wird. Unter Verwendung dieser Annahmen kann das elektrostatische Potential innerhalb und außerhalb des gelösten Moleküls mit Hilfe der linearisierten Poisson-Boltzmann-Gleichung berechnet werden:

$$\nabla[\epsilon(\vec{r})\nabla\varphi(\vec{r})] - \bar{\kappa}^2\varphi(\vec{r}) + 4\pi\rho(\vec{r}) = 0 \quad (3.8)$$

wobei

$$\bar{\kappa} = \sqrt{\epsilon} \cdot \kappa = \sqrt{\frac{8\pi e^2 N_A I}{k_B T}} = \sqrt{\frac{8\pi F^2 I}{RT}}$$

mit

$\vec{r}$ : Ortsvektor

$\epsilon(\vec{r})$ : Dielektrizitätskonstante

$\varphi(\vec{r})$ : Elektrostatisches Potential

$\rho(\vec{r})$ : Ladungsdichte des gelösten Moleküls

$\kappa$ : Inverse Debye-Hückel-Länge

$I$ : Ionenstärke

$N_A$ : Avogadrokonstante

$e$ : Elementarladung

$k_B$ : Boltzmann-Konstante

$T$ : Temperatur

$F$ : Faraday-Konstante

$R$ : Gaskonstante

Eine analytische Lösung der Poisson-Boltzmann-Gleichung ist nur für sehr einfache Systeme wie Kugeln oder Ebenen möglich. Deshalb wird das elektrostatische Potential von Molekülen über ein Finite-Differenzen-Verfahren ermittelt [89-93].

Das Coulomb-Gesetz beschreibt das elektrostatische Potential in einem homogenen Medium mit einem über den gesamten Raum konstanten Wert der Dielektrizitätskonstanten:

$$\varphi(\vec{r}) = \frac{1}{4\pi\epsilon} \cdot \sum_{i=1}^N \frac{q_i}{d_i(\vec{r})} \quad (3.9)$$

mit

$N$ : Anzahl der Atome im Molekül

$q_i$ : Partialladung des Atoms  $i$

$d_i(\vec{r})$ : Abstand des Atoms  $i$  zum Ort der Potentialberechnung  $\vec{r}$

Um die Langreichweitigkeit der elektrostatischen Wechselwirkungen und die Polarisierbarkeit der Lösungsmittels auch im Coulomb-Ansatz zu berücksichtigen, wurde von Brooks et al.[94] eine Modifikation der einfachen Abstandsfunktion des Coulomb-Gesetzes vorgestellt. Der verwendete *Shifted-Force*-Ansatz skaliert das elektrostatische Potential, so daß es bis zu einem *Cutoff*-Radius  $d_0$  auf Null abgesunken ist:

$$\varphi(\vec{r}) = \begin{cases} \frac{1}{4\pi\epsilon} \cdot \sum_{i=1}^N \frac{q_i}{d_i} \cdot \left[ 1 - \left( \frac{d_i}{d_0} \right)^2 \right]^2 & \text{für } d_i \leq d_0 \\ 0 & \text{für } d_i > d_0 \end{cases} \quad (3.10)$$

Dadurch ist es möglich, die explizite Einbeziehung der Einflüsse des Lösungsmittels mit dem analytisch berechenbaren Coulomb-Gesetz zu kombinieren.

### 3.4 Lokale Lipophilie/Hydrophilie

Hydrophobe Effekte spielen bei zahlreichen biochemischen Vorgängen, so auch bei der Proteinfaltung und der molekularen Erkennung, eine entscheidende Bedeutung. Das in dieser Arbeit verwendete Verfahren zur Quantifizierung und Lokalisierung der molekularen Hydrophobie beruht auf dem Verteilungskoeffizient  $P$ , der das Verteilungsgleichgewicht einer Substanz zwischen Wasser und einem wasserunlöslichen organischen Lösungsmittel beschreibt. Richards et al. [95] konnten zeigen, daß die Freie Enthalpie des Transfers und somit auch der Logarithmus des Verteilungskoeffizienten  $\log P$  als Funktion der dem Lösungsmittel zugänglichen Oberfläche darstellbar ist. Aufbauend auf diesem Konzept wurde ein Verfahren entwickelt, daß die Freie Transferenthalpie durch ein Oberflächenintegral über eine Freie Oberflächenenergiedichte (*Free Energy Surface Density*) als Funktion des Abstandes eines Molekülklements zur Oberfläche ermittelt [96].

Die zur Berechnung der Oberflächenenergiedichte herangezogenen Fragmentwerte beruhen auf dem von Ghose und Crippen [97] konzipierten und von Viswanadhan et al. [98,99] erweiterten atomaren Fragmentsystem. Für die räumliche Projektion der Fragmentbeiträge auf die molekulare Oberfläche wurden in der Literatur verschiedene Ansätze vorgeschlagen [100-103]. Bei der in dieser Arbeit verwendeten Abstandsfunktion (Heiden-Moeckel-Funktion) [103] tragen abgeschirmte, im Molekülinneren liegende Atome nur einen geringen Wert zur Hydrophobie bei, während exponierte Fragmente entsprechend höher gewichtet werden:

$$g_{HM}(d; c, \delta) = \frac{e^{-2c/\delta} + 1}{e^{2(d-c)/\delta} + 1} \quad (3.11)$$

mit

$d$ : Abstand des Inkrements zur Oberfläche

$c$ : *Cutoff*-Parameter

$\delta$ : Parameter zur Beeinflussung der Steigung

Die unter Verwendung dieser Abstandsfunktion berechneten Werte auf der Oberfläche werden anschließend normiert. Dadurch wird erreicht, daß die Werte auf der Oberfläche in der Größenordnung der atomaren Fragmentwerte liegen [103]. Die lokale Lipophilie/Hydrophilie kann somit an jedem Oberflächenpunkt wie folgt berechnet werden:

$$L_{HM} = \frac{\sum_{i=0}^N f_i \cdot g_{HM}(d_i; c, \delta)}{\sum_{i=0}^N g_{HM}(d_i; c, \delta)} \quad (3.12)$$

mit

$L_{HM}$ : Lokale Lipophilie/Hydrophilie auf der Oberfläche

$f_i$ : Fragmentwerte nach Ghose und Crippen

$N$ : Anzahl der Atome im Molekül

$g_{HM}$ : Heiden-Moeckel-Funktion (Gleichung 3.11)

### 3.5 Fähigkeit zur Ausbildung von Wasserstoffbrückenbindungen

Wasserstoffbrückenbindungen treten zwischen einem elektronegativen Atom mit kovalent gebundenem Wasserstoff (Donor) und einem weiteren elektronegativen Atom mit einem freien Elektronenpaar (Akzeptor) auf. In biochemischen Systemen treten als Akzeptoratome Sauerstoff, Stickstoff, Schwefel und eventuell Phosphor auf, als Donorgruppen Sauerstoff, Stickstoff und Schwefel mit mindestens einem daran gebundenen Wasserstoffatom. Aufgrund der geometrischen Restriktivität der Wasserstoffbrücken sind sie wesentlich an der exakten Ausrichtung der Moleküle in der Komplexstruktur beteiligt.

Heiden [65] stellte ein einfaches Verfahren vor, um die Fähigkeit einzelner Atome zur Bildung von Wasserstoffbrücken auf der molekularen Oberfläche darzustellen. Jedem Oberflächenpunkt wird der Wert der Fähigkeit zur Wasserstoffbrückenbindung des nächstgelegenen Atoms zugeordnet. Dabei entspricht ein Wert von -1 einem Wasserstoffakzeptor, +1 einem Wasserstoffdonor und 0 keiner Wasserstoffbrücken-Aktivität.



Durch das so auf der Oberfläche erzeugte Muster der Akzeptor / Donorpositionen können qualitative Informationen über die Bildung von Wasserstoffbrücken in Komplexen erhalten werden. Quantitative Aussagen können allerdings damit nicht getroffen werden.

### **3.6 Fuzzifizierung der topographischen und physikochemischen Eigenschaften**

Für die weitere Charakterisierung der topographischen und physikochemischen Eigenschaften auf der Oberfläche sollen diese linguistischen Variablen zugeordnet werden. Voraussetzung dafür ist eine adäquate Umsetzung der verschiedenen Eigenschaften in *Fuzzy*-Mengen, deren Zugehörigkeitsfunktionen für den nachfolgenden Vergleich der Eigenschaften der Bindungsregionen optimiert werden müssen. Heiden et al. [63,65] entwickelten ein System von *Fuzzy*-Mengen zur Beschreibung der linguistischen Variablen Topographie, Elektrostatik und Lipophilie, mit deren Hilfe molekulare Oberflächen in zusammenhängende Domänen unterteilt werden. Dieses System wird in dieser Arbeit auf die oben beschriebenen topographischen und physikochemischen Eigenschaften übertragen und für den *Docking*-Algorithmus optimiert.

Für die Optimierung der Zugehörigkeitsfunktionen wurde das im folgenden beschriebene Verfahren entwickelt, das auf dem Vergleich der topographischen und physikochemischen Eigenschaften in der Binderegion der Komplexe beruht. Dazu werden diese Eigenschaften zuerst auf einer gemeinsamen Oberfläche berechnet. Als gemeinsame Oberfläche wird die *Separating Surface* von Keil et al. [104], die als virtuelle Kontaktfläche im halben Abstand zwischen den Komplexpartnern definiert ist, verwendet. Da die topographischen Eigenschaften nur am Ort der molekularen Oberfläche definiert sind, müssen diese über einen Projektionsalgorithmus auf der *Separating Surface* berechnet werden. Für beide Moleküle werden auf einem dreidimensionalen Gitter alle Werte der Oberflächenpunkte der molekularen Oberfläche abstandsgewichtet bis zu einem Abstand von 3 Å gemittelt. Als Abstandsfunktion wird eine einfache lineare Funktion verwendet. Die Funktionswerte auf der *Separating Surface* werden anschließend durch trilineare Interpolation der Gitterpunkte bestimmt und den korrespondierenden linguistischen Variablen zugeordnet. Die zu dieser linguistischen Variablen gehörenden Zugehörigkeitsfunktionen werden so lange optimiert, bis die über die gesamte *Separating Surface* ermittelte Komplementarität am größten ist. Als Quantifizierung der Komplementarität wird die *Dissimilarity*-Funktion (siehe Gleichung 2.9) verwendet.

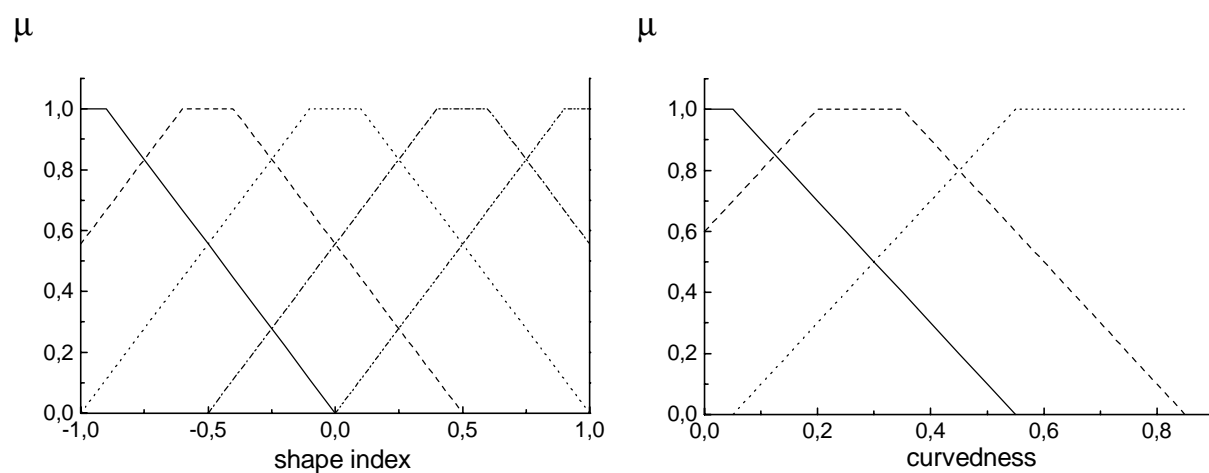


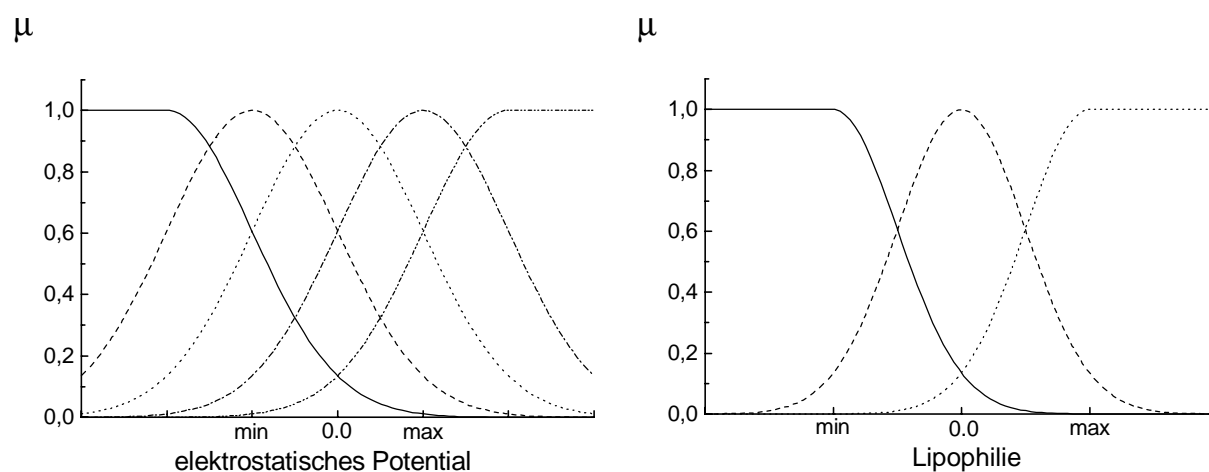
Abb. 3.5:

a) Zugehörigkeitsfunktionen der linguistischen Variable Topographie  $\mathcal{L}_{topo}$

— Loch  
 - - Spalt  
 . . . . Sattel  
 - . . . Grat  
 - - - - Pfropf

b) Zugehörigkeitsfunktionen der linguistischen Variable Krümmung  $\mathcal{L}_{curv}$

— schwache Krümmung  
 - - mittlere Krümmung  
 . . . . starke Krümmung



c) Zugehörigkeitsfunktionen der linguistischen Variable Elektrostatik  $\mathcal{L}_{esp}$

— stark negativ  
 - - negativ  
 . . . . neutral  
 - . . . positiv  
 - - - - stark positiv

d) Zugehörigkeitsfunktionen der linguistischen Variable Lipophilie  $\mathcal{L}_{lipo}$

— hydrophil  
 - - neutral  
 . . . . lipophil

Zur Beschreibung der Topographie der Oberfläche wurden zwei linguistische Variablen  $\mathcal{L}_{topo}$  und  $\mathcal{L}_{curv}$  definiert. Die erste Variable Topographie  $\mathcal{L}_{topo}$  beschreibt die Form der Oberfläche und wird über den *Shape Index* als Basisvariable bestimmt. Sie ist ähnlich wie die von Heiden vorgestellte, auf dem *STI* aufbauende linguistische Variable durch die Klassen „Loch“, „Spalt“, „Sattel“, „Grat“ und „Pfropf“ gegeben (Abb. 3.5a). Die Neigung der Oberfläche wird durch die auf der *Curvedness* beruhende, zweite linguistische Variable Krümmung  $\mathcal{L}_{curv}$  beschrieben (Abb. 3.5b). Diese wird in drei Klassen („schwache“, „mittlere“ und „starke Krümmung“) unterteilt. Durch Kombination der beiden die Topographie beschreibenden linguistischen Variablen ergibt sich somit ein mathematischer Ausdruck für die verbale Beschreibung der Oberfläche (wie z.B. tiefes Loch entspricht Loch mit starker Krümmung).

Die linguistische Variable Elektrostatik  $\mathcal{L}_{esp}$  (Abb. 3.5c) wird in die Klassen „stark negativ“, „negativ“, „neutral“, „positiv“ und „stark positiv“ eingeteilt. Im Gegensatz zu Heiden [63,65] sind die *Fuzzy*-Mengen aber nicht durch stückweise lineare Funktionen, sondern durch Gauß-Funktionen definiert. Die Variable  $\mathcal{L}_{esp}$  wird für alle Berechnungsmethoden des elektrostatischen Potentials verwendet. Die vierte zur Charakterisierung der Oberfläche verwendete linguistische Variable Lipophilie  $\mathcal{L}_{lipo}$  wird aus den drei gaußförmigen Klassen „hydrophil“, „neutral“ und „lipophil“ gebildet (Abb. 3.5d). Für eine gleichmäßige Verteilung der Klassen von  $\mathcal{L}_{esp}$  und  $\mathcal{L}_{lipo}$  über den gesamten Wertebereich der entsprechenden Basisvariablen werden die beiden Parameter *min* und *max* eingeführt. Im Fall des elektrostatischen Potentials wurden diese Parameter auf  $\pm 33\%$  und für die Lipophilie auf  $\pm 50\%$  des betragsmäßig größten Werts der Eigenschaft auf der molekularen Oberfläche gesetzt.

## Kapitel 4

### Algorithmische Konzepte zur Lösung des *Docking*-Problems

In diesem Kapitel werden die zur Behandlung des *Docking*-Problems verwendeten algorithmischen Konzepte vorgestellt. Durch Kombination dieser Algorithmen entsteht ausgehend von den in der *Brookhaven Protein Data Bank* (PDB) [105] gespeicherten Atomkoordinaten der Komplexpartner eine Methode zur Vorhersage der Struktur von Protein-Ligand- und Protein-Protein-Komplexen.

Der Algorithmus besteht aus fünf unabhängigen Teilschritten (Abb. 4.1):

1. Einteilung der molekularen Oberflächen
2. Ermittlung von Komplexstrukturen anhand komplementärer Oberflächeneigenschaften
3. Zusammenfassen ähnlicher Komplexstrukturen zu einer mittleren Struktur mit Hilfe eines Clusteralgorithmus (*Clustering*)
4. Energieoptimierung der Komplexstrukturen unter Verwendung starrer Molekülstrukturen
5. Optimierung unter Berücksichtigung der Ligandenflexibilität mit einem genetischen Algorithmus.

In Kapitel 4.1 wird die Unterteilung der molekularen Oberflächen in Bereiche mit ähnlichen topographischen bzw. physikochemischen Eigenschaften, im weiteren als Domänen bezeichnet, beschrieben. Die so entstehenden Domänen der beiden Komplexpartner werden mit dem in Kapitel 4.2 beschriebenen Algorithmus verglichen und erste Vorschläge für mögliche Komplexstrukturen gemacht. Diese Komplexstrukturen können mit den Teilschritten 3 bis 5 in beliebiger Reihenfolge weiterverarbeitet werden. So können ähnliche Strukturen mit Hilfe eines Clusteralgorithmus zu einer mittleren Komplexstruktur zusammengefaßt werden. Dies wird im weiteren als *Clustering* bezeichnet und ist in Kapitel 4.3 beschrieben.

Daneben können die Komplexstrukturen mit der *Downhill Simplex*-Methode oder mit einem genetischen Algorithmus optimiert werden. Durch den *Downhill Simplex*-Algorithmus (Kapitel 4.5) wird der Ligand unter Verwendung starrer Molekülstrukturen ins nächstgelegene lokale Minimum verschoben. Dadurch eignet sich der *Downhill Simplex*-Algorithmus zur Voroptimierung einer Vielzahl möglicher Komplexstrukturen. Der in Kapitel 4.6 vorgestellte genetische Algorithmus berücksichtigt dagegen die Flexibilität des Liganden und soll vor allem für die flexible Optimierung der energetisch günstigsten Komplexstruktur dienen. Die

für die Energieminimierungsverfahren eingesetzte Zielfunktion wird in Kapitel 4.4 beschrieben.

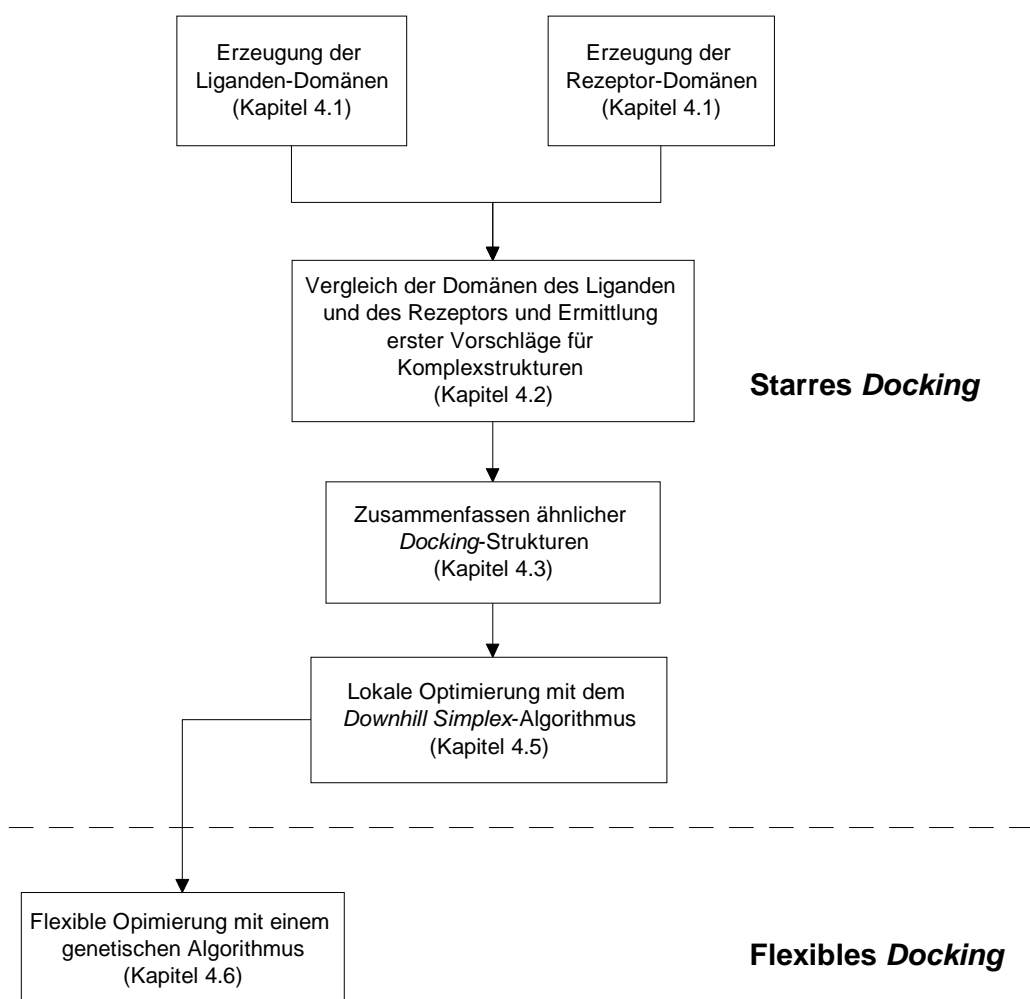


Abb. 4.1: Schematische Darstellung des *Docking*-Algorithmus

## 4.1 Segmentierung der molekularen Oberflächen

Bei der Berechnung der *Solvent Accessible Surface* ergeben sich entsprechend den kanonischen Krümmungen drei verschiedene Arten von Oberflächenbereichen: konkave und konvexe Bereiche sowie Sattellbereiche [74]. Lin et al. [25] entwickelten ein Verfahren, das jeden dieser Oberflächenbereiche durch einen repräsentativen Oberflächenpunkt (*Sparse Critical Point*) sowie einen mittleren Normalenvektor charakterisiert. Die so entstehende Einteilung der molekularen Oberfläche hat jedoch den Nachteil, daß nur sehr kleine Oberflächenbereiche beschrieben werden und somit auf jeder Oberfläche eine große Anzahl von Referenzpunkten entstehen. Für die weitere Verarbeitung muß die Anzahl der Referenzpunkte weiter verringert werden, indem sehr dicht beieinander liegende Bereiche zu

einem kombiniert werden und für jeden Komplexpartner nur eine bestimmte Art von Oberflächenbereichen, z.B. konkave Bereiche des Rezeptors und konvexe Bereiche des Liganden, für den Vergleich herangezogen werden [28].

Heiden et al. [63,65] entwickelten aufbauend auf der *Fuzzifizierung* des *STI*, des elektrostatischen Potentials und der Lipophilie (siehe Kapitel 3.6) eine Einteilung der molekularen Oberfläche in größere, zusammenhängende Oberflächenbereiche. Dazu werden für alle Oberflächenbereiche die Zugehörigkeit zur entsprechenden linguistischen Variablen berechnet und benachbarte Punkte mit ähnlichen Eigenschaften zu Domänen zusammengefaßt. Bei dem von Heiden vorgeschlagenen Algorithmus entstehen aber Domänen, die sich in ihrer Größe und Form sehr stark unterscheiden und darum nur schwer miteinander verglichen werden können. Deswegen wird im folgenden ein Segmentierungsalgorithmus vorgestellt, der die Ausdehnung auf größere Oberflächenbereiche ermöglicht und trotzdem leicht zu vergleichende Oberflächendomänen erzeugt.

Der Algorithmus beruht auf dem Prinzip des Wachstums der Domäne ausgehend von kritischen Punkten bis zu einer durch den Benutzer definierten Maximalgröße. Die kritischen Punkte werden als lokale Maxima der Zugehörigkeit zu einer Klasse der linguistischen Variablen bestimmt. Die Domänen werden dann der Klasse mit dem größten Zugehörigkeitswert des kritischen Punktes zugeordnet. Für die Topographie entstehen so z.B. Domänen der Klassen *Loch*, *Spalt*, *Sattel*, *Grat* und *Pfropf* (siehe Kapitel 3.6). Im weiteren wird dies als Zugehörigkeit einer Domäne zu einer bestimmten Klasse bezeichnet. Im Gegensatz zur von Heiden et al. [63,65] vorgeschlagenen Segmentierung können Oberflächenpunkte mehreren Domänen zugeordnet werden. Es kommt also zu einer Überlappung der Domänen. Die Grenzen der Domänen werden nicht scharf definiert, jeder Oberflächenpunkt besitzt vielmehr eine Zugehörigkeit zu jeder Domäne. So entsteht für jede Domäne ein kleiner Kernbereich mit Oberflächenpunkten mit großer Zugehörigkeit zur entsprechenden Domäne. Zur Vergrößerung der Domäne können immer Punkte mit entsprechend kleinerer Zugehörigkeit ( $\alpha$ -Niveaumenge mit kleinem Wert  $\alpha$ ) mit berücksichtigt werden. Die Maximalgröße der Domäne ist schließlich erreicht, wenn kein Oberflächenpunkt mehr eine Zugehörigkeit über Null zur entsprechenden Domäne besitzt.

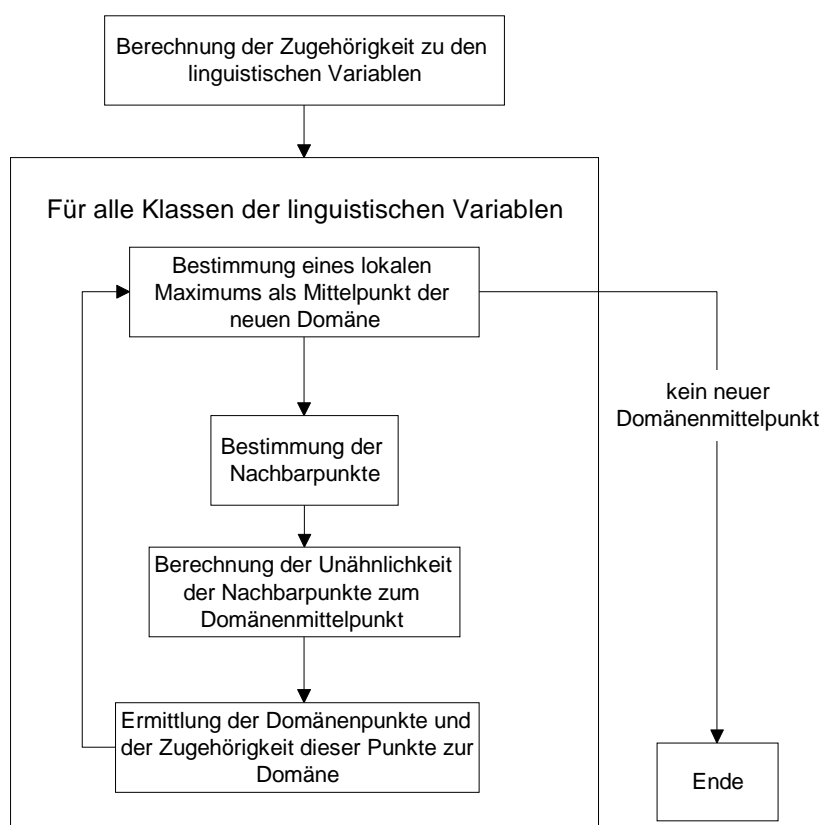


Abb. 4.2: Schematische Darstellung der Einteilung der molekularen Oberfläche in Domänen

Der erste Schritt der Segmentierung ist die Berechnung der Zugehörigkeit zu den Klassen der linguistischen Variablen der entsprechenden Eigenschaft. Als Basisvariable können dabei der *Shape Index*, das elektrostatische Potential sowie die Lipophilie verwendet werden. Die einzelnen Klassen der linguistischen Variablen werden nacheinander zur Segmentierung herangezogen, indem die Oberflächenpunkte nach steigender Zugehörigkeit zur untersuchten Klasse der linguistischen Variablen sortiert werden. Der Oberflächenpunkt mit der größten Zugehörigkeit, im weiteren als Zentralpunkt bezeichnet, wird als Mittelpunkt der neuen Domäne gewählt. Die Ähnlichkeit bzw. Unähnlichkeit aller Oberflächenpunkte zum Zentralpunkt werden nach Gleichung 2.9 berechnet. Befinden sich in einer Fläche von einem Drittel der Maximalgröße der Domäne um den Zentralpunkt Oberflächenpunkte, die eine Ähnlichkeit von  $S_{LV}(A, B) \geq 0.9$  besitzen, wird der Schwerpunkt dieser Punkte und des Zentralpunkts bestimmt. Aus diesem Schwerpunkt wird ein neuer Domänenmittelpunkt berechnet. Dieser ergibt sich als der Oberflächenpunkt, der einer Projektion des Schwerpunkts auf die Oberfläche nächstgelegenen ist. Dadurch ist gewährleistet, daß sich der neue Domänenmittelpunkt im Zentrum eines lokalen Extremwerts der untersuchten Eigenschaft befindet.

Ausgehend von diesem Domänenmittelpunkt werden alle Nachbarpunkte entlang der Oberfläche bestimmt, bis der durch diese Punkte definierte Oberflächenbereich die Maximalgröße einer Domäne erreicht hat. Für die Ermittlung der Nachbarpunkte wird der für die Berechnung der globalen Krümmungen beschriebene Algorithmus verwendet (siehe Kapitel 3.2). Zu jedem dieser Nachbarpunkte wird der kürzeste Weg entlang der Dreieckskanten bis zum Domänenmittelpunkt ermittelt. Eine Domäne wird dann aus allen Punkten gebildet, deren Ähnlichkeit zum Zentralpunkt größer als 0.5 ist ( $\alpha$ -Niveaumenge mit  $\alpha = 0.5$ ) und auf deren Verbindungsweg zum Domänenmittelpunkt kein Oberflächenpunkt liegt, dessen Ähnlichkeit unter 0.5 liegt. Ist dabei die Größe der Domäne kleiner als eine bestimmte Mindestgröße, wird die Domäne bei der weiteren Rechnung nicht mehr berücksichtigt.

Jedem Oberflächenpunkt einer Domäne wird ein Zugehörigkeitswert zugeordnet, der sich aus dem Abstand zum Domänenmittelpunkt berechnet.

$$\mu_{dom,i} = \frac{(d'_{max} - d_i)}{d'_{max}} \quad (4.1)$$

wobei

$$d'_{max} = d_{max} + 0.1 \text{ \AA}$$

mit

$\mu_{dom,i}$ : Domänenzugehörigkeit der Punktes  $i$

$d_i$ : Abstand des Oberflächenpunktes  $i$  zum Domänenmittelpunkt [ $\text{\AA}$ ]

$d_{max}$ : Maximaler Abstand eines Oberflächenpunktes der Domäne zum Domänenmittelpunkt [ $\text{\AA}$ ]

Für alle Punkte der Domäne ergibt sich eine Zugehörigkeit größer Null, für alle anderen Punkte gleich Null. Die so generierten Domänen haben eine ungefähr kreisförmige Gestalt. Damit werden die Domänen miteinander vergleichbar. Die unterschiedliche Größe der Domänen auf gegenüberliegenden Teilen der beiden Komplexpartner kann durch die unscharfe Definition der Domänengrenze aufeinander angepaßt werden. So können beim Vergleich von einer großen Domäne nur Oberflächenpunkte mit einer großen Domänenzugehörigkeit herangezogen werden, während für kleine Domänen eine  $\alpha$ -Niveaumenge der Domänenzugehörigkeit mit kleinem Wert  $\alpha$  verwendet wird.

Der Segmentierungsalgorithmus wird dann mit dem Oberflächenpunkt mit der nächst größten Zugehörigkeit zur untersuchten Klasse, der noch keiner Domäne der selben Klasse zugeordnet ist, fortgesetzt. Als Zentralpunkte für die Domäne einer bestimmten Klasse sollen aber nur Oberflächenpunkte wirken, deren Zugehörigkeit zu dieser Klasse größer ist als zu allen



anderen Klassen. Anderenfalls würde eine Domäne durch einen Oberflächenpunkt definiert, dessen Eigenschaften den der Klasse der Domäne zugeordneten Kriterien nur wenig oder gar nicht entsprechen. Ist kein Oberflächenpunkt mehr vorhanden, der eine Klassenzugehörigkeit über einem gewissen Minimalwert aufweist, wird mit der nächsten Klasse der linguistischen Variablen fortgefahren. Der Minimalwert richtet sich nach der zur Segmentierung herangezogenen Eigenschaft und beträgt für den *Shape Index* 0.85 und für das elektrostatische Potential und die Lipophilie 0.9 bzw. 0.7. Diese Werte entsprechen ungefähr den Überschneidungspunkten der Zugehörigkeitsfunktionen der Klassen der linguistischen Variablen (siehe Abb. 3.5) und gewährleisten, dass jeder Oberflächenpunkt nur als Zentralpunkt einer Domäne der Klasse, zu der seine Zugehörigkeit am größten ist, gewählt werden kann.

## 4.2 Vergleich der Domänen der molekularen Oberflächen

Die molekularen Oberflächen beider Komplexpartner werden nach dem oben beschriebenen Verfahren einer Segmentierung anhand des *Shape Index*, des elektrostatischen Potentials sowie der Lipophilie unterzogen. Durch einen Vergleich der Eigenschaften der so generierten Domänen können Vorschläge für eine mögliche Komplexstruktur ermittelt werden. Der Programmablauf des Domänenvergleichs ist in Abbildung 4.3 schematisch dargestellt. Die Einzelschritte werden in den nachfolgenden Kapiteln beschrieben.

### 4.2.1 Bestimmung der Eigenschaften einer Domäne

Die topographischen bzw. physikochemischen Eigenschaften einer Domäne werden durch den Mittelwert der entsprechenden Eigenschaften der Oberflächenpunkte definiert. Um eine im späteren Algorithmus nötige Anpassung der Größe der Domänen zu ermöglichen, werden diese gemittelten Eigenschaften nicht nur für die gesamte Domäne, sondern auch für verschiedene  $\alpha$ -Niveaumengen der Domänenzugehörigkeit berechnet. Dazu wird der Durchschnitt aller Oberflächenpunkte, die zur entsprechenden  $\alpha$ -Niveaumenge gehören, gebildet. Neben der Stütze der Domäne (strenge  $\alpha$ -Niveaumenge mit  $\alpha = 0$ ), also der gesamten Domäne, werden auch die  $\alpha$ -Niveaumengen mit Werten  $\alpha = 0.1$  bis  $0.9$  mit einem Inkrement von  $0.1$  verwendet. Zu jeder dieser  $\alpha$ -Niveaumengen wird zusätzlich die Anzahl der enthaltenen Oberflächenpunkte sowie die Größe der daraus resultierenden Oberfläche bestimmt. Die Beschreibung der Lage der Domäne erfolgt durch einen repräsentativen Punkt sowie einem Normalenvektor. Der repräsentative Punkt ergibt sich als Schwerpunkt der Oberflächenpunkte eines Kernbereiches der Domäne. Als Kernbereich werden dabei alle

Punkte mit einer Domänenzugehörigkeit größer als 0.9 angesehen. Der Normalenvektor berechnet sich als Mittelwert der Normalenvektoren der Punkte des Kernbereichs.

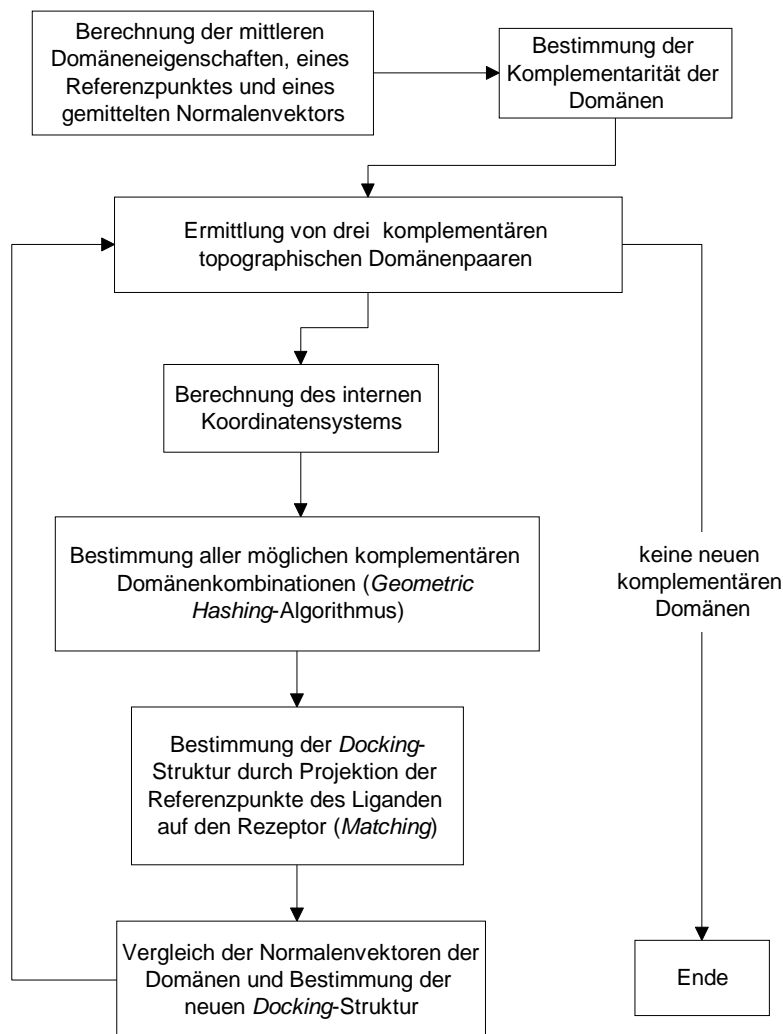


Abb. 4.3: Schematische Darstellung des Vergleichs der in Domänen eingeteilten molekularen Oberflächen des Rezeptors und des Liganden

#### 4.2.2 Vergleich der Domäneneigenschaften

Für die Ausbildung eines stabilen Komplexes ist wie in Kapitel 2.1 beschrieben eine Komplementarität der molekularen Oberflächen nötig. Ist eine Komplexbildung gegeben, sollte sich diese durch die Existenz komplementärer Oberflächendomänen beschreiben lassen. Der hier vorgestellte Algorithmus quantifiziert mit Hilfe eines Vergleichs der Domäneneigenschaften die Komplementarität jeder möglichen Kombination einer Domäne des Rezeptors mit einer Domäne des Liganden und ermittelt daraus mögliche Bindungsregionen auf den molekularen Oberflächen.

Zuerst wird ein negatives Bild des Rezeptors erstellt, indem zu jeder Domäne  $A$  eine ideal komplementäre Domäne  $A'$  definiert wird. Die Eigenschaften der Domäne  $A'$  ergeben sich nach folgendem Schema: Die Größe und die Anzahl der Oberflächenpunkte von  $A'$  entsprechen der Originaldomäne. Für topographische Domänen ergeben sich die Werte des *Shape Index* von  $A'$  als negativer Wert von  $A$ . Dies entspricht einer Zuordnung einer konvexen Domäne zu einer konkaven Domäne mit dem gleichen Verhältnis der beiden globalen Krümmungen und umgekehrt. Ein Sattelpunkt wird wieder einem Sattelpunkt zugeordnet. Um gleiche Absolutbeträge der Krümmungen von  $A'$  und  $A$  zu gewährleisten, erhält die komplementäre Domäne den Originalwert der *Curvedness*. Domänen des negativen Bilds des Rezeptors, die nach dem elektrostatischen Potential bzw. der Lipophilie eingeteilt wurden, erhalten den negativen Wert des gemittelten elektrostatischen Potentials bzw. den Originalwert der Lipophilie von  $A$ .

Für den weiteren Vergleich werden für alle Kombinationen aus Rezeptor-Domänen  $A$  und Ligand-Domänen  $B$  entsprechende  $\alpha$ -Niveaumengen bestimmt, deren Größen sich am ehesten entsprechen. Dadurch wird eine Anpassung der Domänengröße möglich, und es können nicht nur die über die gesamte Domäne gemittelten Eigenschaften, sondern auch ein vom Kernbereich bis zur Größe der kleineren Domäne gehendes Profil der Eigenschaften verglichen werden. Für die zum Vergleich herangezogenen  $\alpha$ -Niveaumengen wird die im Kapitel 3.6 vorgestellte Zugehörigkeit zu den entsprechenden linguistischen Variablen berechnet, durch deren Vergleich anschließend die Komplementarität der Domänen quantifiziert wird. Dazu werden im Falle der topographischen Eigenschaften die Unähnlichkeit  $D_{LV}(A', B)$  der linguistischen Variablen Topographie und Krümmung mit Hilfe der *Dissimilarity*-Funktion (Gleichung 2.9) ermittelt. Die Komplementarität der einzelnen linguistischen Variablen  $C_{LV}(A, B)$  ergibt sich dann als Mittelwert des Komplements der Unähnlichkeit der Topographie bzw. Krümmung der verschiedenen Kombinationen von  $\alpha$ -Niveaumengen (Gleichung 4.2) und die Gesamtkomplementarität der Domänen als Mittelwert von  $C_{LV}(A, B)$  der Topographie und der Krümmung:

$$C_{LV}(A, B) = \frac{1}{N} \cdot \sum_{i=0}^N 1 - D_{LV,i}(A', B) \quad (4.2)$$

mit

$A$ : Domäne des Rezeptors

$A'$ : Ideal komplementäre Domäne zu  $A$

$B$ : Domäne des Liganden

$N$ : Anzahl der Kombinationen von  $\alpha$ -Niveaumengen ähnlicher Größe

$D_{LV,i}(A', B)$ : Unähnlichkeit der  $i$ -ten Kombination von  $\alpha$ -Niveaumengen

$C_{LV}(A, B)$ : Komplementarität von  $A$  und  $B$

Die Komplementarität des elektrostatischen Potentials und der Lipophilie können nicht auf die gleiche Weise berechnet werden. Während bei der Topographie eine komplementäre Oberfläche durch die ideal komplementäre Domäne beschrieben werden kann, sind bei diesen Eigenschaften eine Vielzahl unterschiedlicher Definitionen mit idealer Komplementarität möglich. Domänen des elektrostatischen Potentials müssen dabei unterschiedliche Vorzeichen, Domänen der Lipophilie gleiches Vorzeichen aufweisen. Unterschiedliche Absolutwerte der Eigenschaften verändern nur die Stärke der resultierenden Wechselwirkungen, verhindern aber nicht die Komplexbildung. Deswegen wird allen Kombinationen der  $\alpha$ -Niveaumengen des elektrostatischen Potentials mit unterschiedlichem Vorzeichen und der Lipophilie mit gleichem Vorzeichen ein Wert der Komplementarität von  $C_{LV}(A, B) = 1.0$  zugeordnet. Für die Berechnung der Komplementarität von Domänenkombinationen mit gleichem Vorzeichen des elektrostatischen Potentials bzw. verschiedenem Vorzeichen der Lipophilie wird auf die Klasse *neutral* der entsprechenden linguistischen Variablen zurückgegriffen. Dazu wird der kleinere Wert der Zugehörigkeiten der Domäne des Liganden und des Rezeptors zur Klasse *neutral* bestimmt und dieser als Komplementarität der Domänen verwendet. Die Gesamtkomplementarität der Domänen ergibt sich wiederum als Mittelwert der Komplementarität der einzelnen Kombinationen der  $\alpha$ -Niveaumengen (Gleichung 4.2).

#### 4.2.3 *Geometric Hashing*-Algorithmus

Eine stabile Komplexstruktur sollte sich durch eine Vielzahl von komplementären Domänenkombinationen, deren geometrische Lage eine gleichzeitige Überlappung der Oberflächen dieser Domänen ermöglicht, auszeichnen. Der zur Auffindung dieser Domänenkombinationen verwendete Algorithmus, im weiteren als *Geometric Hashing*-Algorithmus bezeichnet, wurde von Schwartz et al. [26,27] für die Robotersteuerung entwickelt. Die Gruppe von Nussinov und Wolfson [106-110] verwendete diesen Algorithmus für eine chemische Fragestellung, wobei anhand der Lage der  $C_\alpha$ -Atome der Aminosäuren ein Strukturvergleich von Proteinen mit geringer Sequenzhomologie vorgenommen wurde. Die Erweiterung auf die von Lin et al. [25] vorgeschlagene Unterteilung der molekularen Oberfläche führte zu einem erfolgreichen *Docking*-Algorithmus [111,112].

Der *Geometric Hashing*-Algorithmus basiert auf dem Vergleich der Anordnung von indizierten Punkten im dreidimensionalen Raum und besteht aus zwei Teilschritten [26,27]: Zum Vergleich werden nur die Koordinaten der Punkte sowie ein den Punkten zugeordneter Index, der die Eigenschaften des Punktes beschreibt, herangezogen. Im ersten Schritt (*Preprocessing*) wird für jedes Molekül ein von der Transformation unabhängiges internes Koordinatensystem definiert. Für jeden Referenzpunkt werden dann die neuen Koordinaten in

diesem System berechnet. Im zweiten Schritt (*Recognition*) wird durch Vergleich der internen Koordinaten die größte Anzahl möglicher komplementärer Domänenkombinationen bestimmt.

In der hier beschriebenen Implementierung sind die Punkte durch die repräsentativen Punkte der Domänen gegeben. Dabei werden die nach den drei Eigenschaften *Shape Index*, elektrostatisches Potential und Lipophilie eingeteilten Domänen verwendet. Die Indizierung der Punkte erfolgt durch die Komplementarität der entsprechenden Domäne zu allen Domänen des anderen Komplexpartners. Weitere indizierte Punkte beschreiben die mögliche Ausbildung von Wasserstoffbrückenbindungen. Dazu wird die auf die Oberfläche projizierte Fähigkeit zur Ausbildung von Wasserstoffbrückenbindungen verwendet. Zu jedem Akzeptor oder Donor wird der Schwerpunkt aller zu diesem Atom gehörenden Oberflächenpunkte berechnet und als Referenzpunkt der Wasserstoffbrückenbindungsfähigkeit verwendet.

Für die Bestimmung des internen Koordinatensystems werden drei nicht colineare Referenzpunkte benötigt. Dazu werden drei Paare topographisch eingeteilter, sich überlappender Domänen ermittelt, deren Komplementarität über  $C_{LV}(A, B) = 0.9$  liegt. Diese drei Domänenkombinationen beschreiben ein lokales Gebiet sterischer Komplementarität der beiden Moleküle. Die repräsentativen Punkte der zu diesen Kombinationen gehörenden Domänen bilden jeweils ein Dreieck für den Rezeptor und den Liganden. Eine gleichzeitige Überlappung der drei repräsentativen Punkte ist nur möglich, wenn die korrespondierenden Seiten der beiden Dreiecke gleich lang sind. Deswegen werden zur Bestimmung des internen Koordinatensystems nur Dreiecke herangezogen, bei denen die Differenzen der Längen der Dreiecksseiten unter einem Maximalbetrag von 1.50 Å liegen. Die Einheitsvektoren dieses neuen Koordinatensystems ergeben sich dann aus den drei repräsentativen Punkten nach Gleichung 4.3:

$$\begin{aligned}\vec{e}_x &= \vec{p}_2 - \vec{p}_1 \\ \vec{e}_y &= \vec{e}_x \times (\vec{p}_3 - \vec{p}_1) \\ \vec{e}_z &= \vec{e}_x \times \vec{e}_y\end{aligned}\tag{4.3}$$

mit

$\vec{e}_x, \vec{e}_y, \vec{e}_z$ : Einheitsvektoren des internen Koordinatensystems

$\vec{p}_1, \vec{p}_2, \vec{p}_3$ : Koordinaten der repräsentativen Punkte der drei Domänen

Für alle anderen repräsentativen Punkte des Rezeptors und des Liganden werden danach die neuen Koordinaten in dem jeweiligen Koordinatensystem berechnet.

Ausgehend von dem komplementären Kernbereich kann im anschließenden *Recognition*-Schritt die Komplexbindungsregion durch das Anfügen weiterer komplementärer Domänen mit korrespondierender Lage in den internen Koordinatensystemen weiter vergrößert werden. Dazu werden repräsentative Punkte des Liganden und des Rezeptors gesucht, deren Koordinaten sich in den internen Systemen nur leicht unterscheiden und für die sich eine Komplementarität von  $C_{LV}(A,B) \geq 0.85$  berechnet. Dabei dürfen sich die Koordinaten der Punkte, die nach der Topographie, dem elektrostatischen Potential und der Lipophilie eingeteilte Domänen repräsentieren, um 2.00 Å in jeder Richtung unterscheiden. Für Punkte der Wasserstoffbrückenbindungsfähigkeit ist eine maximale Differenz von 1.00 Å in jede Richtung erlaubt. Dadurch wird berücksichtigt, daß Wasserstoffbrückenbindungen im Gegensatz zu den anderen Eigenschaften eine größere geometrische Spezifität aufweisen.

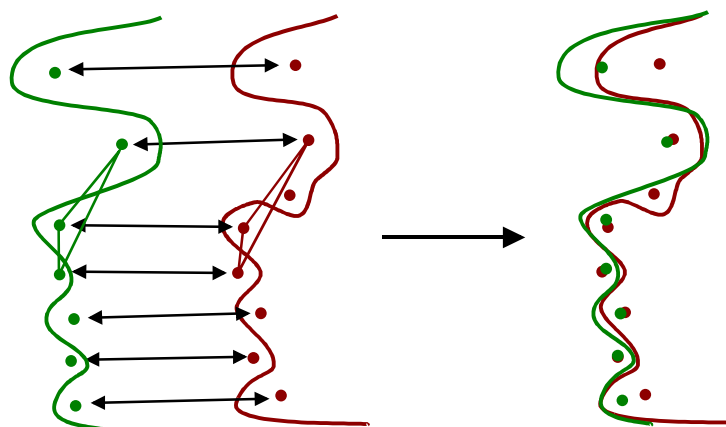


Abb. 4.4: Schematische Darstellung der Bestimmung der Komplexstruktur  
Die repräsentativen Punkte des Liganden (grün) und des Rezeptors (rot) der komplementären Domänenkombinationen werden durch den *Least Squares Fit*-Algorithmus aufeinander projiziert. Die zwei Dreiecke bestimmen die für die Berechnung der internen Koordinatensysteme verwendeten Domänen.

Mit den so bestimmten Domänenkombinationen kann nun ein Vorschlag für eine Komplexstruktur bestimmt werden. Dazu werden alle Domänenkombinationen, die die im *Recognition*-Schritt beschriebenen Komplementaritäts- und Abstandskriterien erfüllen, in eine Liste aufgenommen. Die repräsentativen Punkte dieser Domänen werden mit Hilfe des Ferro-Hermans-Algorithmus [113] aufeinander projiziert, der durch eine analytische Minimierung der Quadrate der Abstände (*Least Squares Fit*, Gleichung 4.4) eine Transformationsmatrize des Liganden berechnet. Über die so ermittelte Transformationsmatrix werden anschließend die Koordinaten der repräsentativen Punkte und die Richtung der gemittelten Normalenvektoren neu berechnet:

$$F = \sum_{i=0}^N \left| \vec{u}_i - \hat{M} \cdot \vec{v}_i \right|^2 \quad (4.4)$$

mit

$F$ : Zielfunktion

$\vec{u}_i, \vec{v}_i$ : Koordinaten der komplementären Domänenpaare

$N$ : Anzahl der Paare repräsentativer Punkte

$\hat{M}$ : Transformationsmatrix des Liganden

Da Domänenkombinationen, deren Normalenvektoren nicht in entgegengesetzte Richtung zeigen, nicht zur Oberflächenüberlappung beitragen können, müssen diese aus der Liste komplementärer Domänen entfernt werden. Dazu wird der Winkel berechnet, der von den Normalenvektoren der beiden Domänen eingeschlossen wird. Ist dieser Winkel kleiner als  $120^\circ$ , sind die beiden Oberflächenbereiche so ausgerichtet, daß es zu keiner ausreichenden Überlappung kommen kann. Diese Domänenkombinationen werden aus der Liste gestrichen und mit dem *Least Squares Fit*-Algorithmus für die verbleibenden Kombinationen eine neue Komplexstruktur berechnet. Dabei kann aber wiederum eine Komplexstruktur entstehen, bei der das Winkelkriterium nicht für alle Domänenkombinationen erfüllt ist. Deswegen wird die Prozedur so lange fortgesetzt, bis keine Domänenkombination mehr dem Winkelkriterium widerspricht. Alle Komplexstrukturen mit zehn und mehr komplementären Domänenpaaren werden für die weitere Verarbeitung gespeichert.

#### 4.2.4 Test auf sterische Überlappung

Bei den mit dem *Geometric Hashing*-Algorithmus bestimmten Komplexstrukturen handelt es sich durch die unscharfe Behandlung der Oberflächen um vorläufige Orientierungen, bei denen es zu einer starken Volumenüberlappung kommen kann. Durch die weiter unten beschriebene Energieoptimierung und die Berücksichtigung der Flexibilität können diese Volumenüberlappungen beseitigt werden. Da diese Optimierungsprozeduren aber sehr zeitaufwendig sind, muß vor allem bei Protein-Protein-Komplexen die Anzahl der zu optimierenden Komplexstrukturen verringert werden. Deswegen werden schon in diesem Stadium des Algorithmus Strukturen verworfen, die relativ zur Größe der Bindungsregion eine starke Volumenüberlappung aufweisen. Als Maß für die sterische Überlappung  $B_{steric}$  wird der in Gleichung 4.5 beschriebene Quotient aus der Größe der Volumenüberlappung und der Größe der Bindungsregion verwendet. Die Volumenüberlappung wird durch die Anzahl der Ligandenatome bestimmt, deren Atomzentrum sich innerhalb des van-der-Waals-Radius eines Rezeptoratoms befindet. Die Größe der Bindungsregion wird durch die Anzahl aller

Ligand-Rezeptor-Atompaare, deren Abstand kleiner als die um 1.0 Å vergrößerte Summe der entsprechenden Atomradien ist, charakterisiert. So wird berücksichtigt, daß es bei größeren Bindungsregionen trotz guter sterischer Komplementarität zu einer größeren Volumenüberlappung kommen kann. Wird ein Wert von 0.04 überschritten, wird die Überlappung als zu groß angesehen und die Komplexstruktur entfernt.

$$B_{steric} = \frac{\sum_{i=1}^N \delta_i}{\sum_{i=1}^N \sum_{j=1}^M \delta_{ij}} \quad (4.5)$$

wobei:

$$\delta_i = \begin{cases} 1 & \text{wenn } \exists d_{ij} < r_j \\ 0 & \text{sonst} \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & \text{wenn } d_{ij} < r_j + r_i + 1.0 \text{ Å} \\ 0 & \text{sonst} \end{cases}$$

mit

$d_{ij}$ : Abstand zwischen Ligandatom  $i$  und Rezeptoratom  $j$  [Å]

$r_i$ : Van-der-Waals-Radius des Ligandatoms  $i$  [Å]

$r_j$ : Van-der-Waals-Radius des Rezeptoratoms  $j$  [Å]

### 4.3 Clusteralgorithmus zur Bestimmung mittlerer Orientierungen aus ähnlichen Komplexstrukturen

Clusteralgorithmen werden zur Einteilung einer Menge von Objekten in Gruppen mit ähnlichen Eigenschaften verwendet. Dazu wird ein Abstandsmaß  $d$  definiert, das die Ähnlichkeit bzw. Unähnlichkeit der Eigenschaften beschreibt. Die verschiedenen in der Literatur vorgestellten Algorithmen lassen sich im wesentlichen in zwei Gruppen unterteilen [114,115]: Heuristische Klassifizierungsverfahren teilen die untersuchten Objekte in eine vordefinierte Anzahl disjunkter Gruppen ein. Hierarchische Verfahren dagegen teilen die Objektmenge anhand des Distanzmaßes sukzessiv in immer größere oder kleinere Cluster ein. Dabei können ausgehend von jedem Objekt in einem eigenen Cluster die ähnlichsten Cluster zu einem größeren kombiniert werden (agglomerative Verfahren) oder ausgehend von der gesamten Objektmenge eines Clusters eine Aufteilung in zwei Cluster, die die größtmögliche



Unähnlichkeit besitzen, erfolgen (divisive Verfahren). Hierarchische Verfahren besitzen den Vorteil, daß die Anzahl der Cluster im vorhinein nicht bekannt sein muß. Sie ergibt sich automatisch, indem der Algorithmus abbricht, wenn der kleinste bzw. größte Abstand zwischen allen Clusterpaaren einen gewissen Wert überschreitet (agglomerative Verfahren) bzw. unterschreitet (divisive Verfahren).

Die Unterschiede in den verschiedenen agglomerativen Verfahren ergeben sich in der Berechnung des Abstandmaßes des neuen Clusters. Häufig verwendete Berechnungsverfahren sind das *Single Linkage*-, *Complete Linkage*-, *Average Linkage*- sowie das Ward-Verfahren. Die Abstände des neuen Clusters  $C_k = C_p \cup C_q$  zu allen anderen Clustern  $C_i$  werden für diese Verfahren wie folgt berechnet:

- *Single Linkage*:  $d_{ki} = \min(d_{pi}, d_{qi})$
- *Complete Linkage*:  $d_{ki} = \max(d_{pi}, d_{qi})$
- *Average Linkage*:  $d_{ki} = \frac{1}{m_p + m_q} \cdot (m_p \cdot d_{pi} + m_q \cdot d_{qi})$
- Ward-Verfahren:  $d_{ki} = \frac{1}{m_p + m_q + m_i} \cdot ((m_p + m_i) \cdot d_{pi} + (m_q + m_i) \cdot d_{qi} - m_i \cdot d_{pq})$

mit

$k$ : Index des neuen Clusters

$p, q$ : Indizes der zu kombinierenden Cluster

$i$ : Index der anderen Cluster

$d_{ij}$ : Abstand zwischen Cluster  $i$  und  $j$

$m_i$ : Anzahl der im Cluster  $i$  enthaltenen Objekte

Im Rahmen der hier beschriebenen Arbeit wurde ein Clusteralgorithmus zur Ermittlung ähnlicher Ligandenorientierungen entwickelt. Diese ähnlichen Orientierungen entstehen, wenn mehrere interne Koordinatensysteme während des *Geometric Hashing*-Algorithmus zu ähnlichen Domänenkombinationen und dadurch zu Transformationsmatrizen mit ähnlichen Orientierungen des Liganden führen. Deswegen kann die Anzahl der weiter zu untersuchenden Komplexstrukturen verringert werden, indem diese ähnlichen Bindungsmodi zu einem Cluster zusammengefaßt und durch eine mittlere Struktur ersetzt werden. Das entwickelte agglomerative, hierarchische Verfahren zeichnet sich bei großen Objektmengen durch eine geringe Laufzeit aus.

Als Distanzmaß wird der *rms*-Abstand (*Root Mean Square Deviation*) zwischen jeweils zwei Orientierungen des Liganden verwendet, der als Wurzel der mittleren quadratischen

Abweichung der Atompositionen zwischen den Strukturen definiert ist. Rarey et al. [116,117] entwickelten ein auf den Transformationsmatrizen beruhendes Verfahren zur schnellen Berechnung der *rms*-Abstände. Dazu werden die Transformationsmatrizen in einen die Rotation und einen die Translation beschreibenden Teil zerlegt:

$$\hat{M} = (\vec{t}, \hat{R}) \quad (4.5)$$

mit  
 $\hat{M} = (m_{ij})$ : Transformationsmatrix des Liganden  
 $\hat{R}$ : Rotationmatrix  
 $\vec{t}$ : Translationsvektor

Damit ist der *rms*-Abstand zweier Ligandenorientierungen 1 und 2 wie folgt definiert:

$$d_{rms}(\hat{M}_1, \hat{M}_2) = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N \left( (\hat{R}_1 \cdot \vec{u}_i + \vec{t}_1) - (\hat{R}_2 \cdot \vec{u}_i + \vec{t}_2) \right)^2} \quad (4.6)$$

mit  
 $\vec{u}_i$ : Atomkoordinaten des Liganden  
 $N$ : Anzahl der Atome des Liganden

Zur effizienten Berechnung des Distanzmaßes kann dies wie folgt transformiert werden [113,118]:

$$d_{rms}(\hat{M}_1, \hat{M}_2) = \sqrt{2 \cdot \sum_{i=1}^3 \sum_{j=1}^3 (\delta_{ij} - a_{ij} - a_{ji}) \cdot X_{ij} + (\vec{t}_1 - \vec{t}_2)^2 + 2 \cdot (\vec{t}_1 - \vec{t}_2)^T \cdot (\hat{R}_1 - \hat{R}_2) \cdot \vec{X}_m} \quad (4.7)$$

mit  
 $\hat{A} = (a_{ij}) = \hat{R}_1^T \cdot \hat{R}_2$   
 $\vec{X}_m = \frac{1}{N} \cdot \sum_{k=1}^N \vec{x}_k$   
 $X_{ij} = \frac{1}{N} \cdot \sum_{k=1}^N x_k(i) \cdot x_k(j)$   
 $\delta_{ij} = \begin{cases} 1 & : \text{falls } i = j \\ 2 & : \text{sonst} \end{cases}$

Jeder Cluster wird durch eine Ligandenorientierung repräsentiert, die einer mittleren Struktur aller im Cluster enthaltenen Komplexstrukturen entspricht. Zur Mittelung der Orientierungen können die Transformationsmatrizen nicht direkt herangezogen werden. Deswegen wird die Rotation jeder Ligandenorientierung mit Hilfe von Quaternionen und die Translation durch die Verschiebung des Molekülschwerpunkts beschrieben. Die mittlere Orientierung ergibt sich

durch die unabhängige Bestimmung einer mittleren Rotation und einer mittleren Translation. Dazu werden die Quaternionen linear interpoliert und ein mittlerer Schwerpunkt berechnet. Aus diesem mittleren Molekülschwerpunkt und der mittleren Rotation wird anschließend die Transformationsmatrix der mittleren Ligandenorientierung ermittelt. Der Abstand zwischen zwei Clustern wird mit Hilfe dieser mittleren Transformationsmatrizen nach Gleichung 4.7 bestimmt. Der Algorithmus kombiniert dann solange die zwei Cluster mit dem kleinsten *rms*-Abstand, bis dieser ein gewisses Maß überschreitet. Als neue Vorschläge für die Komplexstrukturen werden die mittleren Ligandenorientierungen der resultierenden Cluster verwendet.

#### 4.4 Zielfunktion für die Optimierung der Komplexstrukturen

Die weitere Optimierung der Komplexstrukturen mit Hilfe des genetischen Algorithmus bzw. dem *Downhill Simplex*-Algorithmus benötigt eine Zielfunktion, die den energetischen Vergleich verschiedener Orientierungen ermöglicht. In der Literatur wurden mehrere Zielfunktionen für das *Docking*-Problem vorgestellt, die sich stark in den verwendeten Parametern und damit der Komplexität der Berechnung unterscheiden [119-131]. Für eine effektive Optimierung einer Vielzahl von Komplexstrukturen muß eine schnelle Berechnung der relativen Energien möglich sein. Dabei muß aber darauf geachtet werden, daß die kristallographisch bestimmte Komplexstruktur als globales Minimum der Funktion erkannt wird und durch das vereinfachte Kraftfeld keine Artefakte erzeugt werden. Solche Artefakte entstehen, wenn durch die Zielfunktion viele Strukturen ungefähr die gleiche Energie wie das globale Minimum aufweisen und der Optimierungsalgorithmus leicht in einem lokalen Minimum gefangen bleibt. Von Gehlhaar et al. [121] wurde eine Zielfunktion vorgestellt, die alle diese Voraussetzungen erfüllt und so auch schon erfolgreich in *Docking*-Algorithmen eingesetzt werden konnte [44,121,132]. Da es sich hierbei um eine empirische Zielfunktion handelt, deren Werte nicht mit experimentellen Bindungsenergien vergleichbar sind, werden die Werte als dimensionslos betrachtet (Einheit: a.u.).

Die intermolekulare Wechselwirkungsenergie ergibt sich aus einem vereinfachten Lennard-Jones-Potential (Abb. 4.5), im weiteren als sterischer Betrag bezeichnet, und Wasserstoffbrückenbindungstermen und ist als paarweise Summe über alle Schweratome des Liganden und des Rezeptors definiert. Die einzelnen Beträge berechnen sich aus der in Abbildung 4.5 dargestellten, stückweise linearen Funktion. Die Parameter für das verwendete paarweise Potential ergeben sich aus dem Typ der beiden an der Interaktion beteiligten Atome und sind in Tabelle 4.1 aufgelistet. Die unterschiedlichen Parameter bei gleichem Funktionsverlauf für die beiden Arten von Interaktionen berücksichtigen, daß die Bildung

einer Wasserstoffbrückenbindung ein stärkeres Gewicht zur Bindungsenergie hat als eine unpolare Interaktion.

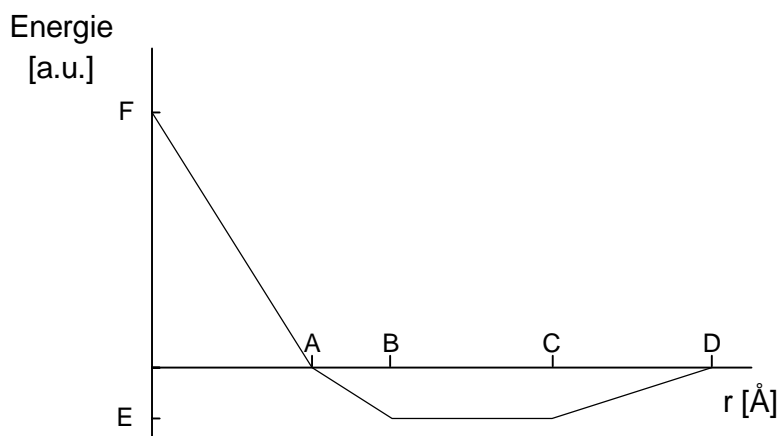


Abb. 4.5: Funktionsverlauf der stückweise linearen Funktion des Ligand-Rezeptor-Paarpotentials

Vier verschiedene Atomtypen werden definiert: Akzeptoren, Donoren, Donoren / Akzeptoren sowie unpolare Atome. Stickstoff- und Sauerstoffatome mit keinen benachbarten Wasserstoffatomen werden als Akzeptoren, Stickstoffatome mit assoziierten Wasserstoffatomen als Donoren und Hydroxy-Sauerstoffatome sowie Wasser als Donor / Akzeptor definiert. Alle anderen Atome werden als unpolar bezeichnet. Jedes Paar eines Liganden- und eines Rezeptoratoms hat genau eine Art der Wechselwirkung. Die Interaktionen zwischen den verschiedenen Atomtypen zeigt Tabelle 4.2.

Tabelle 4.1: Parameter der stückweise linearen Funktion zur Berechnung der Ligand-Rezeptor-Paarpotentiale

Typ der Interaktion	A [Å]	B [Å]	C [Å]	D [Å]	E [a.u.]	F [a.u.]
Sterisch	3.4	3.6	4.5	5.5	-0.4	20.0
Wasserstoffbrücke	2.3	2.6	3.1	3.4	-2.0	20.0

Tabelle 4.2: Typen der paarweise atomaren Interaktionen

Atomtypen	Donor	Akzeptor	Donor / Akzeptor	unpolar
<b>Donor</b>	Sterisch	H-Brücke	H-Brücke	Sterisch
<b>Akzeptor</b>	H-Brücke	Sterisch	H-Brücke	Sterisch
<b>Donor / Akzeptor</b>	H-Brücke	H-Brücke	H-Brücke	Sterisch
<b>Unpolar</b>	Sterisch	Sterisch	Sterisch	Sterisch

Die intramolekulare Energie eines flexiblen oder teilweise flexiblen Moleküls ergibt sich aus einem nichtbindenden Term und einem Torsionswinkelpotential. Dabei werden nur Einfachbindungen zwischen zwei  $sp^3$ -hybridisierten Atomen oder zwischen einem  $sp^3$ -hybridisierten und einem  $sp^2$ -hybridisierten Atom als frei drehbar behandelt. Alle anderen Einfachbindungen sowie Einfachbindungen in Ringen und Mehrfachbindungen werden als starr betrachtet. Alle Bindungslängen und Winkel werden während der Optimierung ebenfalls konstant gehalten. Das Torsionswinkelpotential hat folgende Form:

$$E = A \cdot [1 - \cos(n\Phi - \Phi_0)] \quad (4.8)$$

mit

$$A = 3.0 \text{ a.u.}, n = 3, \Phi_0 = \pi$$

für Bindungen zwischen zwei  $sp^3$ -hybridisierten Atomen

$$A = 1.5 \text{ a.u.}, n = 2, \Phi_0 = 0$$

für Bindungen zwischen einem  $sp^3$ -hybridisierten Atom und einem  $sp^2$ -hybridisierten Atom

Durch einen zusätzlichen Term werden bei der Optimierung intramolekulare Überlappungen im Liganden und Rezeptor vermieden. So wird die nichtbindende, intramolekulare Energie um einen konstanten Wert von 10000 a.u. erhöht, wenn der Abstand zwischen zwei Atomen des gleichen Moleküls kleiner als 2.35 Å wird. Neben den so berechneten inter- und intramolekularen Wechselwirkungen werden keine weiteren Beträge, wie z.B. elektrostatische Wechselwirkungen oder entropische Effekte, in der Zielfunktion berücksichtigt.

Für die Berechnung der Bindungsenergie werden fünf verschiedene Modelle benutzt, die sich nur im Grad der Berücksichtigung der molekularen Flexibilität unterscheiden:

Das erste und einfachste Modell behandelt den Rezeptor sowie den Liganden ohne Berücksichtigung der inneren Freiheitsgrade. Dieses Modell kann vor allem für eine schnelle Voroptimierung einer Vielzahl von Komplexstrukturen verwendet werden.

Die nächsten zwei Ansätze unterscheiden sich im Grad der Flexibilität des Liganden. Für kleine Liganden können alle rotierbaren Einfachbindungen als flexibel angesehen werden. Dies wird im weiteren als vollflexibles Modell bezeichnet. Bei größeren Liganden hätte die Behandlung aller Einfachbindungen eine sehr große, nicht handhabbare Zahl an flexiblen Bindungen zur Folge. Deswegen wird ein teilflexibles Verfahren verwendet. Da es sich bei diesen Liganden ausschließlich um Proteine handelt, werden nur die Einfachbindungen der Seitenketten der an der Bindung beteiligten Aminosäuren als flexibel behandelt. Als an der Bindung beteiligt werden alle Aminosäuren des Liganden angesehen, die maximal 3.0 Å vom Rezeptor entfernt sind. Die Konformation der Seitenketten, die in das Molekülinnere gerichtet sind, wird stark durch das Grundgerüst des Proteins bestimmt. Dadurch ist bei einer Konformationsänderung dieser Seitenketten eine Änderung des gesamten oder von Teilen des Grundgerüsts nötig. Da solche globalen Strukturänderungen durch den Algorithmus nicht erfaßt werden können, werden die Aminosäuren, die nicht zur Oberfläche des Moleküls beitragen, als starr betrachtet. Dadurch wird gewährleistet, daß das Grundgerüst des Proteins bestehen bleibt und nur diejenigen Seitenketten optimiert werden, die auch wirklich einen Beitrag zur Bindungsenergie liefern.

Die letzten beiden Verfahren betrachten die Seitenketten des Rezeptors im Bindungsbereich als flexibel und unterscheiden sich wieder im Grad der Flexibilität des Liganden. Die rotierbaren Bindungen der Rezeptorseitenketten werden auf die gleiche Weise bestimmt wie im teilflexiblen Verfahren des Liganden. Für den Liganden kann wiederum das voll- oder teilflexible Verfahren verwendet werden.

Zur schnelleren Berechnung der Bindungsenergien werden für alle Modelle nur die Atome berücksichtigt, die einen Abstand von unter 10.0 Å zum jeweils anderen Molekül haben. Im Fall von Proteinen werden zusätzlich alle Atome der entsprechenden Aminosäuren in die Berechnung mit einbezogen. Bei allen Modellen mit starrgehaltenem Rezeptor können die intermolekularen Wechselwirkungen durch einen gitterbasierten Algorithmus sehr effizient berechnet werden. Dazu werden für jeden der vier Atomtypen die nichtbindenden Wechselwirkungen des Rezeptors auf einem Gitter mit einem Gitterabstand von 0.5 Å berechnet. Dies entspricht einem Wechselwirkungspotential, das durch den Rezeptor an jedem Ort der Raumes erzeugt wird. Die Wechselwirkungsenergie zwischen Ligand und Rezeptor ergibt sich dann durch trilineare Interpolation der acht benachbarten Gitterpunkte um die jeweilige Atomposition des Ligandatoms. Die intermolekulare Energie wird folglich als

Summe über alle Atome des Liganden ermittelt und nicht über alle Paare von Ligand- und Rezeptoratomen. Dieses Verfahren reduziert die benötigte Rechenzeit sehr stark.

## 4.5 *Downhill Simplex*-Minimierer

Die *Downhill Simplex*-Methode wurde von Nelder und Mead [68] zur Optimierung mehrdimensionaler Probleme vorgeschlagen. Der Algorithmus benötigt keine explizite Berechnung von Ableitungen der Zielfunktion und eignet sich deswegen sehr gut in Kombination mit der oben vorgestellten Energiefunktion. Ein *Simplex* ist eine geometrische Figur, die bei einem N-dimensionalen Problem aus N+1 Punkten besteht. Im zweidimensionalen Fall ist ein *Simplex* zum Beispiel ein Dreieck, im dreidimensionalen eine trigonale Pyramide.

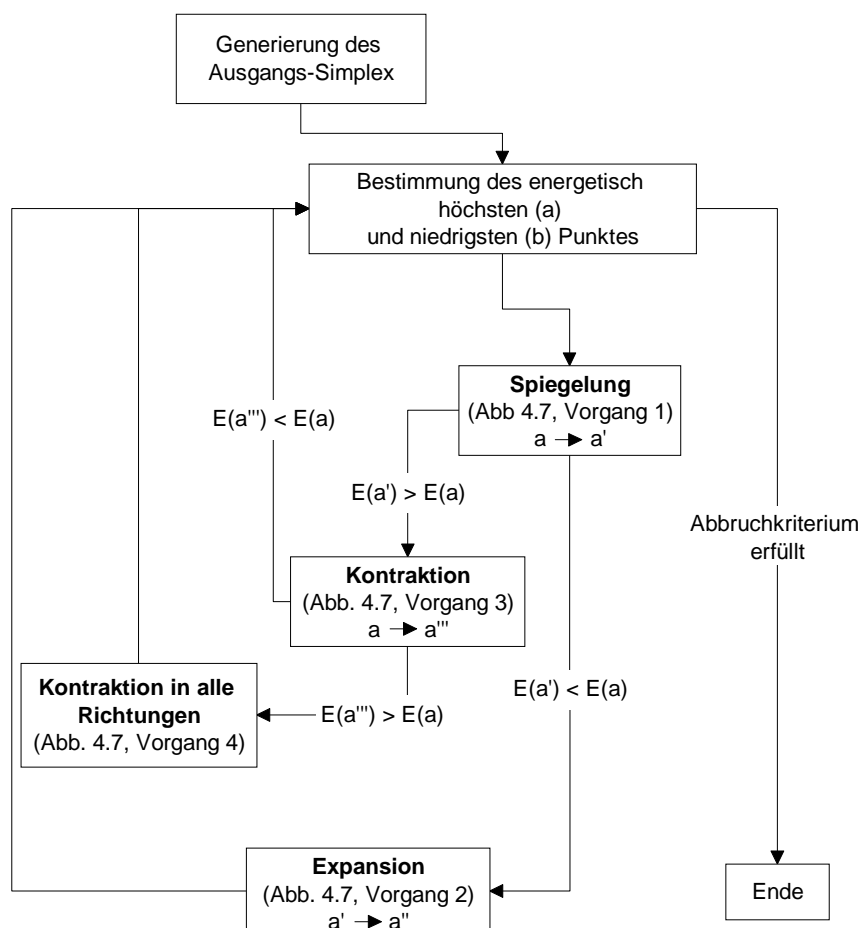


Abb. 4.6: Schematische Darstellung des Programmablaufes des *Downhill Simplex*-Algorithmus (Beschreibung der einzelnen Vorgänge siehe Text und Abb. 4.7)

- |      |                               |       |                              |
|------|-------------------------------|-------|------------------------------|
| a:   | energetisch höchster Punkt    | a':   | gespiegelter Punkt           |
| a'': | neuer Punkt nach Expansion    | a''': | neuer Punkt nach Kontraktion |
| b:   | energetisch niedrigster Punkt |       |                              |

Ausgehend von einem N-dimensionalen *Simplex* versucht der Algorithmus diesen immer weiter zu optimieren, bis ein lokales Minimum erreicht ist. Dabei entsprechen die Dimensionen des *Simplex* den zu optimierenden Parametern. Im Fall des starren *Dockings* werden die Rotation durch die drei Euler-Winkel und die Translation durch die Verschiebungen in x-, y- und z-Richtung beschrieben. Die Koordinaten des ersten Punktes des Ausgangs-*Simplex* sind durch die Parameter der zu optimierenden Ligandenorientierung gegeben. Die Koordinaten der anderen Punkte ergeben sich durch die Addition eines Wertes von  $5.0^\circ$  bzw.  $0.5 \text{ \AA}$  zum entsprechenden Parameter des *Simplex*. So repräsentieren die Punkte des Ausgangs-*Simplex* die Ausgangsorientierung des Liganden und leicht verschobene bzw. rotierte Orientierungen. Wird die Flexibilität des Liganden und des Rezeptors berücksichtigt, wird die Dimension des *Simplex* pro rotierbarer Bindung um 1 erhöht. Für die Initialisierung der rotierbaren Bindungen wird die betreffende Bindung eines Punktes des *Simplex* um  $10^\circ$  gedreht. Für die so definierten Orientierungen werden nach der im vorhergehenden Kapitel beschriebenen Zielfunktion die Energien berechnet.

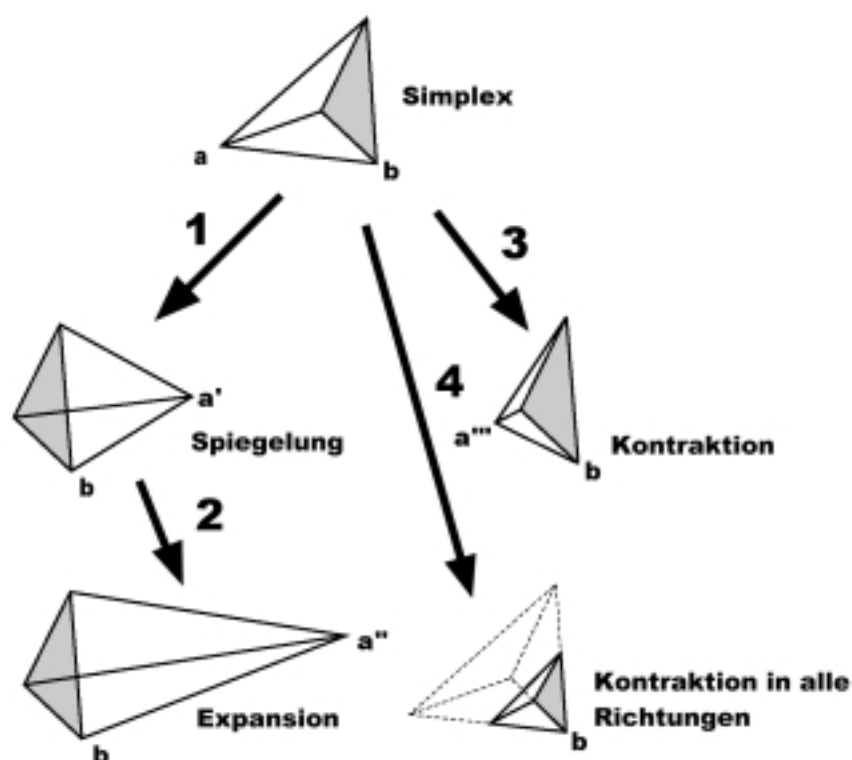


Abb. 4.7: Schematische Darstellung der Operationen des *Downhill Simplex*-Algorithmus für einen *Simplex* mit drei zu optimierenden Variablen (Erklärung siehe Text)

- |      |                               |       |                              |
|------|-------------------------------|-------|------------------------------|
| a:   | energetisch höchster Punkt    | a':   | gespiegelter Punkt           |
| a'': | neuer Punkt nach Expansion    | a''': | neuer Punkt nach Kontraktion |
| b:   | energetisch niedrigster Punkt |       |                              |



Die Optimierung erfolgt dann, indem der jeweils energetisch höchstgelegene Punkt schrittweise in immer niederenergetische Bereiche verschoben wird [133]. Im ersten Schritt wird der energetisch höchstgelegene Punkt am Schwerpunkt aller anderen Punkte des *Simplex* gespiegelt (Abb. 4.7, Vorgang 1). Wird die Energie verringert, wird versucht, durch eine Expansion in diese Richtung eine noch bessere Orientierung zu ermitteln (Expansion, Abb. 4.7, Vorgang 2). Liegt der gespiegelte Punkt energetisch über dem Originalpunkt, wird mit einem Punkt zwischen dem Originalpunkt und dem Schwerpunkt versucht, eine Verbesserung zu erzielen. Dieser Schritt (Abb. 4.7, Vorgang 3) wird als Kontraktion bezeichnet. Kann auch so eine Optimierung nicht erreicht werden, werden alle Punkte des *Simplex* in Richtung des niedrigsten Punktes verschoben (Abb. 4.7, Vorgang 4) und mit dem so entstehenden *Simplex* der Algorithmus fortgesetzt.

Der Algorithmus bricht ab, wenn sich die durch die Punkte repräsentierten Orientierungen ähnlich werden. Als Kriterium dafür gilt, daß sich alle Punkte dem lokalen Optimum soweit angenähert haben, daß die über diese Ähnlichkeit definierte Toleranz unter  $10^{-5}$  fällt. Die Toleranz wird über die Energiewerte des höchsten und des niedrigsten Punktes bestimmt [133]:

$$r_{tol} = \frac{2.0 \cdot |E_{high} - E_{low}|}{|E_{high}| + |E_{low}|} \quad (4.8)$$

mit

$r_{tol}$ : Toleranzwert

$E_{high}$ : Energie des energetisch höchstgelegenen Punktes

$E_{low}$ : Energie des energetisch tiefsten Punktes

Der Algorithmus wird auch beendet, wenn eine maximale Anzahl von Schritten ausgeführt ist und keine Optimierung erreicht werden kann. Bei der starren Optimierung wird die maximale Anzahl auf 500 Schritte gesetzt. Dadurch soll gewährleistet werden, daß nur zur Ausgangsorientierung sehr nah gelegene Minima gefunden werden und die Rechenzeit durch die Optimierung schlechter Ausgangsstrukturen nicht unnötig verlängert wird.

## 4.6 Genetischer Algorithmus zur flexiblen Optimierung von Komplexstrukturen

In Kapitel 2.3 wurde die Theorie der genetischen Algorithmen kurz vorgestellt. Hier wird auf die für den in dieser Arbeit verwendeten genetischen Algorithmus verwendeten Parameter näher eingegangen. Der Ablauf des Algorithmus ist in Abbildung 4.7 schematisch dargestellt. Für die Implementierung des genetischen Algorithmus konnte auf die *Genetic Algorithm Library* von Levine [69] zurückgegriffen werden. Diese Programm-Bibliothek enthält alle für den Aufbau eines genetischen Algorithmus benötigten Unterprogramme.

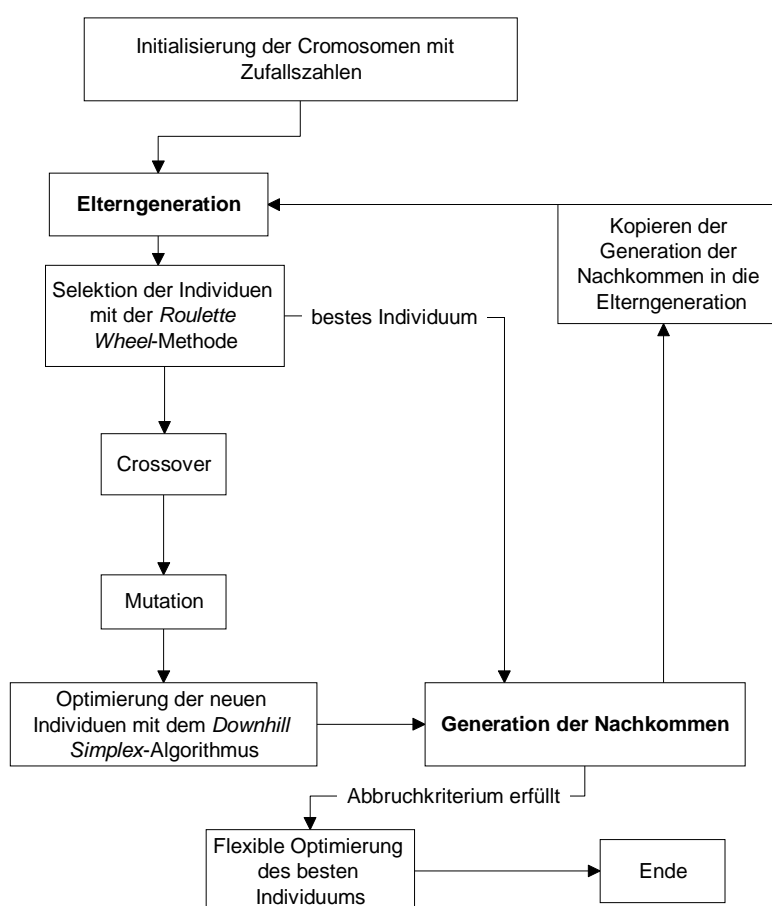


Abb. 4.7: Schematische Darstellung des genetischen Algorithmus

Die Codierung der *Docking*-Variablen erfolgt in einem Vektor aus reellen Zahlen. Dabei repräsentieren die ersten reellen Zahlen die Transformation des starren Liganden. Die nachfolgenden Vektorelemente werden den Torsionswinkeln der einzelnen drehbaren Bindungen zugeordnet. Für kleine Liganden wird das vollflexible Modell, für größere Proteinliganden das teilflexible Modell verwendet. Der Rezeptor wird als starr betrachtet, kann aber auch als teilflexibles Modell in den Algorithmus implementiert werden. Die

gewählte Generationsgröße ist 200 Individuen. Zur Initialisierung werden den Parametern der einzelnen Individuen Zufallszahlen aus bestimmten Intervallen um die Orientierung der Ausgangsstruktur zugeordnet. Für die Translation wird ein Intervall von  $\pm 5.0$  Å in jede Richtung um die Ausgangsstruktur verwendet. Die Rotation des gesamten Liganden wird auf das Intervall von  $\pm 20.0^\circ$  beschränkt. Der Torsionswinkel jeder drehbaren Bindung kann sich auf einen Wert im gesamten Bereich von  $\pm 180.0^\circ$  einstellen.

Die Selektion der für die weiteren Schritte verwendeten Individuen erfolgt mit Hilfe der *Roulette Wheel*-Methode. Dabei wird jedem Individuum eine bestimmte Wahrscheinlichkeit zugeordnet und das zu selektionierende Individuum durch einen Zufallsalgorithmus bestimmt. Dies entspricht einem Roulettespiel, bei dem die einzelnen Zahlen je nach ihrer *Fitness* größere bzw. kleinere Bereiche auf dem Rouletterad zugeordnet bekommen. Die Wahrscheinlichkeit jedes Individuums errechnet sich mit Hilfe der oben beschriebenen Energiefunktion, wobei die Zahlenwerte nachträglich auf ein nicht negatives, monoton ansteigendes Intervall projiziert werden. Das beste Individuum der Elterngeneration wird direkt in die neue Population übernommen. Alle anderen selektierten Individuen werden weiteren Operationen (Crossover und Mutation) unterzogen. Dabei ist es möglich, daß Individuen erst ein Crossover durchlaufen und anschließend noch mutiert werden. Dadurch soll eine schnellere Untersuchung des gesamten Konfigurationsraums ermöglicht werden. Jedes Individuum wird mit 85%iger Wahrscheinlichkeit einem Zwei-Punkt-Crossover unterworfen. Anschließend wird jedes Gen des Individuums mit einer Wahrscheinlichkeit des zweifachen Reziprokwerts der Vektorlänge mutiert. Dies bedeutet, daß in jedem Individuum im Durchschnitt zwei Gene mutiert werden. Dazu wird der Werte des zu mutierenden Gens mit einer Zufallszahl multipliziert. Diese Zufallszahl wird durch eine Gauß-Verteilung bestimmt. Dabei werden für den Mittelwert bzw. die Standardabweichung die in der *Genetic Algorithm Library* implementierten Standardwerte von 0.0 bzw. 0.1 verwendet. Der neue Wert des Gens ergibt sich durch Addition des so berechneten Wertes zum Originalwert. Jeder so neu generierte Nachkomme wird mit Hilfe des *Downhill Simplex*-Minimierers 100 Schritte lokal optimiert. Dabei wird für den Liganden wiederum die Flexibilität der Seitenketten berücksichtigt. Die Schrittzahl des *Downhill Simplex*-Minimierers stellt einen Kompromiß zwischen der bis zur Konvergenz benötigten Anzahl von Generationen und der benötigten Zeit für die Berechnung der *Fitness* jedes einzelnen Individuums dar. Wird die Schrittzahl erhöht, sind weniger Generationen bis zur Konvergenz nötig und die Wahrscheinlichkeit, das globale Minimum zu finden, steigt. Andererseits verläuft die Rechenzeit für die lokale Minimierung jedes Individuums proportional zur Schrittzahl, wodurch die benötigte Rechenleistung für den Generationenwechsel des genetischen Algorithmus stark ansteigt.

Zur Beendigung der iterativen Generierung neuer Individuen wurden zwei Abbruchkriterien benutzt. Der genetische Algorithmus endet, wenn eine maximale Anzahl von Generationen

erzeugt wurde. Diese maximale Anzahl wurde auf 1000 gesetzt, da durch die lokale Optimierung der einzelnen Individuen bereits nach dieser Anzahl von Generationen fast immer Konvergenz erreicht wird. Der Algorithmus endet auch, wenn sich der Energiewert des besten Individuums seit 50 Generationen nicht mehr verbessert. Anschließend wird die Komplexstruktur des Individuum mit der besten *Fitness* mit Hilfe des *Downhill Simplex*-Algorithmus über 1000 Zyklen weiter optimiert. Dazu wird sowohl für den Liganden als auch für den Rezeptor das flexible Modell verwendet.

## Kapitel 5

### Ergebnisse und Diskussion

Strukturen biochemischer Komplexe lassen sich experimentell durch Röntgenstrukturanalyse oder NMR-Spektroskopie ermitteln. Die Atomkoordinaten der so bestimmten Komplexstrukturen sind in der *Brookhaven Protein Data Bank (pdb)* [105] abgelegt. Die *Brookhaven Protein Data Bank* enthält zur Zeit 11000 Einträge biochemischer Moleküle und Molekülkomplexe. Zur Überprüfung der Leistungsfähigkeit des vorgestellten *Docking*-Algorithmus wurden für 35 mit Röntgenstrukturanalyse bestimmte Komplexe mögliche Komplexstrukturen bestimmt und diese mit der experimentellen Struktur aus der Datenbank verglichen. Dabei wurden die Beispiele verschiedenen Arten von biochemischen Komplexen entnommen: 10 Komplexe eines Enzyms mit einem niedermolekularen Inhibitor (Tab. 5.1), 18 Komplexe eines Enzyms mit einem inhibierenden Protein (Tab. 5.2), drei Protein-Dimere (Tab. 5.3) und vier Antikörper-Antigen-Komplexe (Tab. 5.4). In einem Fall eines Antikörper-Antigen-Komplexes (2igf) wurde nur ein Fragment des Antigens verwendet.

Tabelle 5.1: *Pdb*-Einträge der Enzym-Inhibitor-Komplexe mit niedermolekularen Inhibitoren

Nr.	<i>pdb</i> -Datei	Chain	Komplexpartner	Aminosäuren Schweratome	Literatur
1	1hne	E I	Human Leukocyte Elastase Methoxysuccinyl-Ala-Ala-Pro-Ala- Chloromethyl Ketone	218 / 1636 5 / 31	[134]
2	1tlp	E I	Thermolysin Phosphoramidon	316 / 2432 3 / 37	[135]
3	2er9	E I	Endothia Aspartic Proteinase (Endothiapepsin) L363,564	330 / 2389 9 / 65	[136]
4	2est	E I	Elastase Trifluoroacetyl-L-Lysyl-L-Alanyl- P-Trifluoromethylphenylanilide	240 / 1822 4 / 31	[137]
5	3apr	E I	Acid Proteinase (Rhizopuspepsin) Reduced Peptide Inhibitor	325 / 2417 8 / 57	[138]
6	4dfr	A,B MTX	Dihydrofolate Reductase Methotrexate	159 / 1280 1 / 33	[139]
7	4hvp	A,B I	HIV-1 Protease N-Acetyl-Thr-Ile-Nle-Ψ(CH <sub>2</sub> - NH)-Nle-Gln-Arg amide	99 / 1516 6 / 54	[140]
8	5apr	E I	Acid Proteinase (Rhizopuspepsin) Pepstatin-Like Renin Inhibitor	325 / 2417 9 / 58	[141]
9	5sga	E P	Proteinase A Ace-Pro-Ala-Pro-Try	181 / 1259 5 / 35	[142]
10	6tmn	E I	Thermolysin Cbz-Gly <sup>P</sup> -(O)-Leu-Leu	316 / 2432 4 / 32	[143]

Tabelle 5.2: *Pdb*-Einträge der Enzym-Inhibitor-Komplexe mit inhibierenden Proteinen

Nr.	<i>pdb</i> -Datei	Chain	Komplekspartner	Aminosäuren Schweratome	Literatur
11	1acb	E I	$\alpha$ -Chymotrypsin Eglin C	245 / 1769 70 / 522	[144]
12	1cbw	A,B,C D	Bovine Chymotrypsin Basic Pancreatic Trypsin Inhibitor	241 / 1759 58 / 454	[145]
13	1cho	E I	$\alpha$ -Chymotrypsin Turkey Ovomucoid Third Domain	245 / 1750 56 / 400	[146]
14	1cse	E I	Subtilisin Carlsberg Eglin C	274 / 1920 71 / 522	[147]
15	1ppf	E I	Human Leukocyte Elastase Turkey Ovomucoid Third Domain	218 / 1636 56 / 418	[148]
16	1tec	E I	Thermitase Eglin C	279 / 1881 70 / 522	[149]
17	1tgs	Z I	Trypsinogen Porcine Pancreatic Secretory Trypsin Inhibitor	229 / 1646 56 / 416	[150]
18	1tpa	E I	Anhydro-Trypsin Pancreatic Trypsin Inhibitor	223 / 1628 58 / 454	[151]
19	2cpk	E I	c-AMP-Dependent Protein Kinase Peptide Inhibitor	350 / 2665 20 / 157	[152]
20	2kai	A,B I	Kallikrein Bovine Pancreatic Trypsin Inhibitor	232 / 1799 58 / 438	[153]
21	2ptc	E I	$\beta$ -Trypsin Pancreatic Trypsin Inhibitor	223 / 1629 58 / 454	[151]
22	2sec	E I	Subtilisin Carlsberg Genetically-Engineered N-Acetyl Eglin C	274 / 1923 71 / 530	[154]
23	2sic	E I	Subtilisin BPN' Streptomyces Subtilisin Inhibitor	275 / 1938 107 / 764	[155]
24	2sni	E I	Subtilisin Novo Chymotrypsin Inhibitor 2	275 / 1940 83 / 513	[154]
25	2tgp	Z I	Trypsinogen Pancreatic Trypsin Inhibitor	229 / 1629 58 / 454	[151]
26	4cpa	I	Carboxypeptidase A $\alpha$ Potato Carboxypeptidase A Inhibitor	307 / 2443 38 / 285	[156]
27	4sgb	E I	Serine Proteinase B Potato Inhibitor PCI-1	185 / 1310 51 / 380	[157]
28	4tpi	Z I	Trypsinogen Arg <sup>15</sup> -Analogue of Pancreatic Trypsin Inhibitor	229 / 1629 58 / 471	[158]

Tabelle 5.3: *Pdb*-Einträge der Protein-Dimere

Nr.	<i>pdb</i> -Datei	Chain	Komplekspartner	Aminosäuren Schweratome	Literatur
29	1aar	A B	Di-Ubiquitin	76 / 601 76 / 602	[159]
30	2ccy	A B	Cytochrome c'	128 / 976 128 / 976	[160]
31	2utg	A B	Uteroglobulin	70 / 548 70 / 548	[161]

Tabelle 5.4: *Pdb*-Einträge der Antikörper-Antigen-Komplexe

Nr.	<i>pdb</i> -Datei	Chain	Komplexpartner	Aminosäuren Schweratome	Literatur
32	1fdl	L,H	IgG1 Fab Fragment (Anti-Lysozym Antibody D1.3, $\kappa$ )	432 / 3308	[162]
		Y	Lysozym	129 / 1001	
33	3hfl	L,H	IgG1 Fab Fragment (Hy/Hel-5)	427 / 3250	[163]
		Y	Lysozym	129 / 1001	
34	3hfm	L,H	IgG1 Fab Fragment (HyHEL-10)	429 / 3295	[164]
		Y	Lysozym	129 / 1001	
35	2igf	L,H P	Igg1 Fab'Fragment Peptide (Residues 69-87 of Myohemerythrin)	440 / 3378 19 / 58	[165]

## 5.1 Aufarbeitung der kristallographischen Strukturdaten

Für die weitere Verwendung der *pdb*-Daten müssen diese aufgearbeitet werden. Da durch Röntgenstrukturanalyse die Bestimmung von leichten Atomen nur sehr schwer möglich ist, enthalten die kristallographischen Strukturdaten keine oder nur beschränkte Informationen über die Lage der Wasserstoffatome. Der erste Schritt der Aufarbeitung besteht in der Bestimmung der Wasserstoffatompositionen. Anschließend werden die molekularen Oberflächen und die ihnen zugeordneten physikochemischen Eigenschaften berechnet. Die Berechnungen können durch das von Keil entwickelte Programmpaket *BLACKBOX* [166] weitestgehend automatisiert durchgeführt werden.

Für die Behandlung von Proteinen steht das Programm *CHARMM* Version 24 mit einem für Proteinstrukturen parametrisierten Kraftfeld zur Verfügung [94]. Zur Ermittlung der Wasserstoffpositionen in den Proteinstrukturen wurden die Wasserstoffatome zuerst addiert und dann bis zur Konvergenz mit dem internen Kraftfeld energetisch minimiert. Dabei wurden die Schweratome auf den kristallographischen Positionen fixiert und für die Coulomb-Wechselwirkungen ein *Shifted Force*-Ansatz mit einem *Cutoff*-Radius von 11 Å benutzt. Da für nichtpeptidische Inhibitoren (1hne, 1tlp, 2er9, 2est, 4dfr, 5apr, 6tmn) keine Kraftfeldparameter im Programm *CHARMM* enthalten sind, wäre für diese Inhibitoren eine langwierige Parametrisierung notwendig. Da dies für eine schnelle Ermittlung der *Docking*-Strukturen aber zu aufwendig ist, wurden für diese Moleküle die Wasserstoffatome mit Hilfe des *Molecular Modeling Software*-Pakets *SYBYL* [167] addiert.

Die *Solvent Accessible Surface* wurde mit dem Programm *MSMS* erzeugt [81]. Der Radius des Probeteilchens betrug dabei standardmäßig 1.4 Å, die Oberflächen haben eine durchschnittliche Punktdichte von 4 Punkten pro Å<sup>2</sup>. Das elektrostatische Potential nach dem

modifizierten Coulomb-Ansatz, die Lipophilie und die Wasserstoffbrückenbindungsfähigkeit wurden mit dem Programm *QUALEN* berechnet [63,166]. Die Berechnung des elektrostatischen Potentials nach dem Poisson-Boltzmann-Ansatz erfolgte mit dem in dieser Arbeit entwickelten Programm *POISSON* [167]. Die Partialladungen wurden für Proteine aus den Parametern des *CHARMM*-Kraftfeldes entnommen. Für nichtpeptidische Inhibitoren wurden die Partialladungen mit Hilfe des Programmpakets *SYBYL* nach der Methode von Gasteiger-Hückel berechnet [167]. Die benötigten atomaren Lipophilieparameter wurden mit dem Programm *MOLFESD* [168] und die Wasserstoffbrückenbindungsfähigkeit mit *GETHB* bestimmt [166]. Den Abschluß der Aufarbeitung bildeten die Berechnung der globalen Krümmungen, des *Shape Index* und der *Curvedness*. Für die Bestimmung der Referenzpunkte für die globalen Krümmungen wurde ein maximaler Abstand von 4.0 Å für Komplexe mit niedermolekularen Inhibitoren und den Komplex 2igf sowie 5.0 Å für alle anderen Komplexarten verwendet.

## 5.2 Segmentierung der molekularen Oberflächen

Für eine möglichst optimale Einteilung der Oberflächen in Domänen muß darauf geachtet werden, daß die Größe der Domänen nicht zu klein gewählt wird, damit sich durch die Datenreduktion die Möglichkeit eines schnellen, effektiven Oberflächenvergleichs ergibt und die Anzahl der ermittelten Komplexstrukturen nicht zu groß wird. Andererseits werden durch die Generierung von zu großen Domänen die spezifischen Eigenschaften der Bindungsregionen nicht mehr beschrieben, so daß die Ermittlung der Komplexstruktur nicht mehr möglich ist. Durch diese einschränkenden Kriterien wurde die Domänengröße für Komplexe mit niedermolekularen Inhibitoren und für den Antikörper-Antigen-Komplex 2igf auf eine maximale Größe von 50 Å<sup>2</sup> festgelegt. Für alle anderen Komplexe ergab sich durch die Optimierung eine Maximalgröße der Domänen von 100 Å<sup>2</sup>. Die Anzahl der sich aus dieser Einteilung ergebenden Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen sind in Anhang 1 bis 4 zusammengestellt. Für alle Komplexe konnten die so resultierenden Domänen erfolgreich in dem anschließenden *Docking*-Experiment verwendet werden.



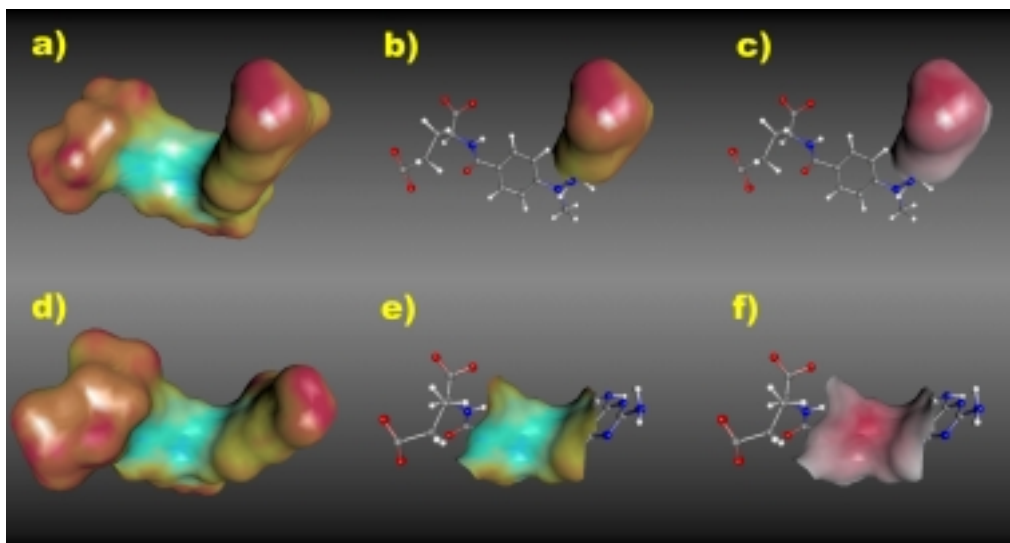


Abb. 5.1: Zwei nach topographischen Kriterien eingeteilte Domänen von Methotrexat. Die Reihe a bis c zeigt einen typischen *Pfropf*, die Reihe d bis f einen *Sattel*  
a und d) gesamte molekulare Oberfläche, farbcodiert nach dem *Shape Index*  
rot: konvexe Bereiche                      blau: konkave Bereiche  
b und e) Domäne der Klasse *Pfropf* bzw. *Sattel*, farbcodiert wie unter a)  
c und f) Domäne der Klasse *Pfropf* bzw. *Sattel*, farbcodiert nach der Zugehörigkeit zur Domäne  
rot: große Zugehörigkeit                      grau: geringe Zugehörigkeit

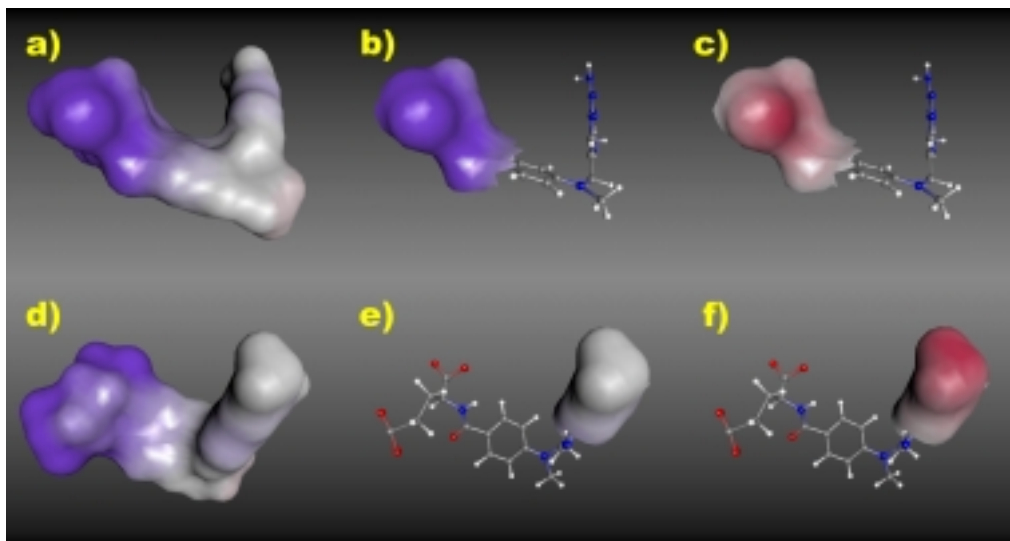


Abb. 5.2: Zwei nach dem elektrostatischen Potential eingeteilte Domänen von Methotrexat. Die Reihe a bis c zeigt eine *stark negativ*, die Reihe d bis f eine *neutrale* Domäne  
a und d) gesamte molekulare Oberfläche, farbcodiert nach dem elektrostatischen Potential  
rot: positive Bereiche                      blau: negative Bereiche  
b und e) Domäne der Klasse *stark negativ* bzw. *neutral*, farbcodiert wie unter a)  
c und f) Domäne der Klasse *stark negativ* bzw. *neutral*, farbcodiert nach der Zugehörigkeit zur Domäne  
rot: große Zugehörigkeit                      grau: geringe Zugehörigkeit

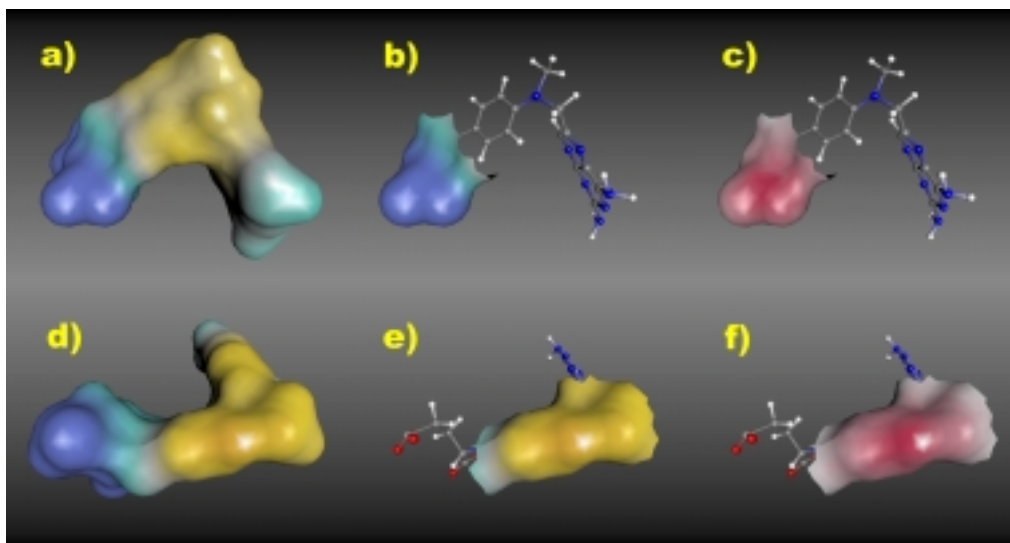


Abb. 5.3: Zwei nach der Lipophilie eingeteilte Domänen von Methotrexat. Die Reihe a bis c zeigt eine *hydrophile*, die Reihe d bis f eine *lipophile* Domäne  
a und d) gesamte molekulare Oberfläche, farbcodiert nach der Lipophilie  
gelb: lipophile Bereiche                      blau: hydrophile Bereiche  
b und e) Domäne der Klasse *hydrophil* bzw. *lipophil*, farbcodiert wie unter a)  
c und f) Domäne der Klasse *hydrophil* bzw. *lipophil*, farbcodiert nach der Zugehörigkeit zur Domäne  
rot: große Zugehörigkeit                      grau: geringe Zugehörigkeit

Die visuelle Überprüfung mit dem Programm *MOLCAD II* [79] zeigt eine gute Wiedergabe der zur Segmentierung herangezogenen Eigenschaften durch die Domänen. In Abbildung 5.1 bis 5.3 sind am Beispiel der molekularen Oberfläche des Methotrexat exemplarisch einzelne Domänen gezeigt. Dabei ist nur der Teil der Domäne gezeigt, der eine Zugehörigkeit zu dieser Domäne von größer als Null aufweist. Dieser Bereich ist einmal nach der entsprechenden Eigenschaft und einmal nach der Zugehörigkeit zur Domäne farbcodiert. Zum Vergleich ist zusätzlich die zur Segmentierung herangezogene Eigenschaft auf der gesamten molekularen Oberfläche dargestellt. Deutlich zu erkennen ist dabei der nahezu kreisförmige Aufbau der Domänen sowie die zur Größenanpassung der Domänen benötigte unscharfe Definition der Domänengrenzen.

### 5.3 Starres *Docking*

Im folgenden Kapitel wird zuerst die Eignung des starren *Docking*-Algorithmus zur Vorhersage von Enzym-Ligand-Komplexen mit niedermolekularen Liganden untersucht. Daran anschließend werden in den Kapiteln 5.3.2 und 5.3.3 die Ergebnisse für einige Komplexe im Detail diskutiert. Die Methode wird in Kapitel 5.3.4 auf die anderen Komplexarten angewendet, und Kapitel 5.3.5 gibt eine allgemeine Bewertung der durch den starren *Docking*-Algorithmus erzielbaren Ergebnisse. Die angegebenen Rechenzeiten beziehen sich, soweit nicht anders angegeben, auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

#### 5.3.1 Vorhersage der Komplexstrukturen niedermolekularer Liganden

Für den Erfolg des Algorithmus ist entscheidend, daß die Oberflächendomänen der Bindungsregionen im ersten Schritt als stark komplementär erkannt werden und somit durch den *Geometric Hashing*-Algorithmus eine oder mehrere Strukturen nahe an der kristallographisch bestimmten Komplexstruktur ermittelt werden können. In Tabelle 5.5 sind die Ergebnisse der Vorhersagen des *Geometric Hashing*-Algorithmus zusammengestellt. Dabei zeigt sich, daß für neun der zehn Komplexe mindestens eine Komplexstruktur mit einem *rms*-Wert unter 2.0 Å ermittelt werden konnte. Die Struktur mit dem geringsten *rms*-Wert zur kristallographischen Struktur wird im weiteren als beste Struktur bezeichnet. Nur für den Komplex aus der humanen Leukocyten-Elastase und Methoxysuccinyl-Ala-Ala-Pro-Ala-Chromomethyl-Ketone (1hne) konnte keine Struktur bestimmt werden, deren Abweichung zum kristallographischen Komplex als ausreichend klein angesehen werden kann. In diesem Fall konnte durch eine Verringerung des Proberadius für die Generierung der Oberfläche auf 1.35 Å die beste Struktur mit einem *rms*-Wert von 1.097 Å bestimmt werden und somit eine Vorhersage der kristallographischen Komplexstruktur erfolgen.

In Tabelle 5.6 sind die Komplexstrukturen mit dem geringsten *rms*-Wert und der größten Anzahl komplementärer Domänenpaare gegenübergestellt. Zusätzlich ist die bestplatzierte Struktur mit einem *rms*-Wert von unter 2.5 Å in die Tabelle aufgenommen. Dabei zeigt sich, daß für jeden Komplex eine Orientierung des Liganden ermittelt werden konnte, die der kristallographischen Komplexstruktur sehr ähnlich ist und die eine große Anzahl komplementärer Domänen ausweist. Durch die unscharfe Definition der *Docking*-Kriterien und den Vergleich der sich über größere Molekülteile erstreckenden Domänen werden auch eine Reihe von Komplexstrukturen ermittelt, die durch die gleiche oder eine größere Anzahl von komplementären Domänenpaaren definiert werden. Diese sollten im Rahmen des

Tabelle 5.5: *Docking*-Vorhersagen der Enzym-Ligand-Komplexe mit niedermolekularen Inhibitoren durch Vergleich der segmentierten Oberflächen (Poisson-Boltzmann-Ansatz)

<i>pdb</i> -Datei	Anzahl komplementärer Domänen <sup>1)</sup>	Anzahl Strukturen	Anzahl von komplementären Domänen der besten Struktur	<i>rms</i> -Wert (Å)	Rechenzeit <sup>2)</sup> (min:s)
<b>1hne</b> <sup>3)</sup>	15676 (1929, 10288, 3459)	790	18	3.270	0:13
<b>1hne</b> <sup>4)</sup>	16264 (2255, 10343, 3666)	1079	23	1.097	0:18
<b>1tlp</b>	18479 (4757, 6304, 3379)	2801	25	0.623	1:01
<b>2er9</b>	31975 (11165, 14369, 6441)	28132	41	0.442	5:59
<b>2est</b>	10226 (2245, 6076, 1905)	497	18	1.875	0:17
<b>3apr</b>	35906 (11413, 18350, 6143)	28699	53	0.918	5:24
<b>4dfr</b>	8388 (2328, 4139, 1921)	1362	28	0.364	0:20
<b>4hvp</b>	23259 (6570, 12430, 4259)	29939	67	0.893	3:42
<b>5apr</b>	24023 (7493, 10533, 5997)	17659	27	1.035	3:50
<b>5sga</b>	11296 (2114, 7078, 2104)	1514	34	0.600	0:19
<b>6tmn</b>	12545 (3650, 6508, 2387)	1516	21	1.057	0:37

1) Anzahl der Domänenpaare mit einer Komplementarität über 0.8

Die erste Zahl ist die Gesamtzahl aller Domänenpaare, in Klammern sind die nach topographischen Kriterien, nach dem elektrostatischen Potential und nach der Lipophilie eingeteilten Domänen getrennt aufgeführt.

Für die Berechnung des elektrostatischen Potentials wurde der Poisson-Boltzmann-Ansatz verwendet.

2) Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor

3) Bei der Generierung der Oberfläche wurde ein Radius der Probekugel von 1.40 Å verwendet.

4) Bei der Generierung der Oberfläche wurde ein Radius der Probekugel von 1.35 Å verwendet.

vorgestellten Modells andere Bindungsmodi der beiden Komplexpartner sein. Durch visuelle Inspektion konnte festgestellt werden, daß die Bindungsregionen dieser Komplexstrukturen komplementäre Oberflächenregionen besitzen und somit als mögliche Komplexstrukturen in Frage kommen. Dies ist exemplarisch am Beispiel des humanen Leukocyten-Elastase / Methoxysuccinyl-Ala-Ala-Pro-Ala-Chromomethyl-Ketone-Komplex (1hne) in Abbildung 5.4 gezeigt.

Tabelle 5.6: Vergleich der Komplexstrukturen mit dem geringsten *rms*-Wert in Bezug auf die kristallographische Struktur, der Strukturen mit den meisten komplementären Domänenpaaren und der Strukturen mit bestem Rang und einem *rms*-Wert unter 2.5 Å (Poisson-Boltzmann-Ansatz)

<i>pdb</i> -Datei	Beste Struktur			Struktur auf Rang 1		Struktur mit bestem Rang und einem <i>rms</i> -Wert unter 2.5 Å		
	Anzahl Domänen	<i>rms</i> -Wert (Å)	Rang	Anzahl Domänen	<i>rms</i> -Wert (Å)	Anzahl Domänen	<i>rms</i> -Wert (Å)	Rang
1hne	23	1.097	55	31	10.496	25	1.891	28
1tlp	25	0.623	6	27	0.954	27	0.954	1
2er9	41	0.442	77	56	16.626	47	1.339	18
2est	18	1.875	9	23	7.535	18	1.875	9
3apr	53	0.918	4	56	1.256	56	1.256	1
4dfr	28	0.364	2	29	0.906	29	0.906	1
4hvp	67	0.893	2	68	1.058	68	1.058	1
5apr	27	1.035	569	46	4.097	37	1.472	38
5sga	34	0.600	2	35	0.685	35	0.685	1
6tmn	21	1.057	61	29	19.940	27	1.755	4

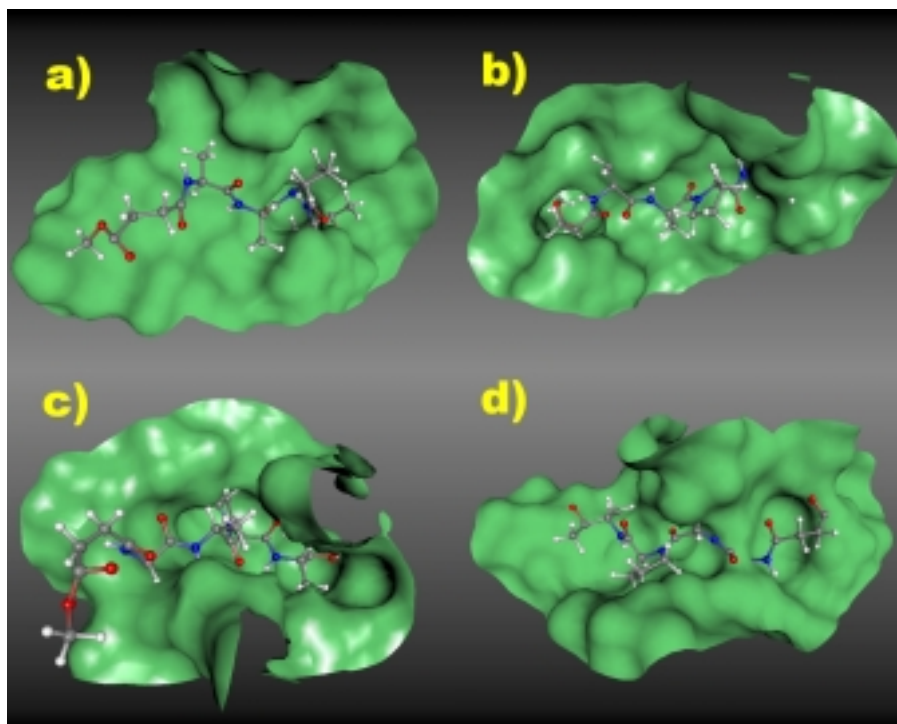


Abb. 5.4: Exemplarische Beispiele von ermittelten Bindungsmodi des humanen Leukocyten Elastase / Methoxysuccinyl-Ala-Ala-Pro-Ala-Chromomethyl-Ketone-Komplexes  
a) beste Struktur  
b) Struktur auf Rang 1  
c) Struktur auf Rang 6  
d) Struktur auf Rang 11

Um den Einfluß der zur Berechnung des elektrostatischen Potentials verwendeten Methode zu untersuchen, wurde der *Docking*-Algorithmus für jeden Komplex ein zweites Mal durchgeführt. In diesem zweiten Lauf wurde für die Berechnung des elektrostatischen Potentials der modifizierte Coulomb-Ansatz verwendet. Die Ergebnisse sind in Tabelle 5.7 und 5.8 zusammengestellt. Aus dem Vergleich mit den Tabellen 5.5 bzw. 5.6 für den Poisson-Boltzmann-Ansatz kann keine Präferenz des *Docking*-Algorithmus für einen der beiden Ansätze zur Berechnung des elektrostatischen Potentials erkannt werden. Beide Ansätze führen zu sehr ähnlichen Ergebnissen, sowohl bei der Anzahl der komplementären Domänenpaare als auch beim *rms*-Wert der besten Struktur. Deswegen wurde im weiteren Verlauf der Arbeit nur der modifizierte Coulomb-Ansatz verwendet, da hiermit das elektrostatische Potential schneller berechnet werden konnte.

Tabelle 5.7: *Docking*-Vorhersagen der Enzym-Ligand-Komplexe mit niedermolekularen Inhibitoren durch Vergleich der segmentierten Oberflächen (modifizierter Coulomb-Ansatz)

<i>pdb</i> -Datei	Anzahl komplementärer Domänen <sup>1)</sup>	Anzahl Strukturen	Anzahl von komplementären Domänen der besten Struktur	<i>rms</i> -Wert (Å)	Rechenzeit <sup>2)</sup> (min:s)
<b>1hne<sup>3)</sup></b>	13172 (1929, 7784, 3459)	590	11	4.912	0:12
<b>1hne<sup>4)</sup></b>	13609 (2255, 7688, 3666)	813	22	1.141	0:16
<b>1tlp</b>	12454 (4755, 4320, 3379)	2421	25	0.929	1:00
<b>2er9</b>	34450 (11164, 16845, 6441)	29962	46	0.583	10:10
<b>2est</b>	7519 (2245, 3369, 1905)	267	13	1.695	0:18
<b>3apr</b>	28927 (11414, 11370, 6143)	23965	53	0.743	5:17
<b>4dfr</b>	6653 (2328, 2404, 1921)	1183	20	0.482	0:19
<b>4hvp</b>	18517 (6570, 7688, 4259)	26768	46	0.804	3:31
<b>5apr</b>	27189 (7490, 13702, 5997)	18792	36	0.635	4:09
<b>5sga</b>	8558 (2114, 4340, 2104)	1141	20	0.791	0:17
<b>6tmn</b>	10379 (3650, 4342, 2387)	1145	17	1.009	0:34

1) Anzahl der Domänenpaare mit einer Komplementarität über 0.8

Die erste Zahl ist die Gesamtzahl aller Domänenpaare, in Klammern sind die nach topographischen Kriterien, nach dem elektrostatischen Potential und nach der Lipophilie eingeteilten Domänen getrennt aufgeführt.

Für die Berechnung des elektrostatischen Potentials wurde der modifizierte Coulomb-Ansatz verwendet.

2) Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor

3) Bei der Generierung der Oberfläche wurde ein Radius der Probekugel von 1.40 Å verwendet.

4) Bei der Generierung der Oberfläche wurde ein Radius der Probekugel von 1.35 Å verwendet.

Tabelle 5.8: Vergleich der Komplexstrukturen mit dem geringsten *rms*-Wert in Bezug auf die kristallographische Struktur, der Strukturen mit den meisten komplementären Domänenpaaren und der Strukturen mit bestem Rang und einem *rms*-Wert unter 2.5 Å (modifizierter Coulomb-Ansatz)

<i>Pdb</i> -Datei	Beste Struktur			Struktur auf Rang 1		Struktur mit bestem Rang und einem <i>rms</i> -Wert unter 2.5 Å		
	Anzahl Domänen	<i>rms</i> -Wert (Å)	Rang	Anzahl Domänen	<i>rms</i> -Wert (Å)	Anzahl Domänen	<i>rms</i> -Wert (Å)	Rang
1hne	22	1.141	22	27	18.388	22	1.141	22
1tlp	25	0.929	3	27	1.376	27	1.376	1
2er9	46	0.583	66	58	1.512	58	1.512	1
2est	13	1.695	30	18	7.422	13	1.695	30
3apr	53	0.743	3	59	0.817	59	0.817	1
4dfr	20	0.482	39	27	19.765	25	0.916	3
4hvp	46	0.804	136	62	1.454	62	1.454	1
5apr	36	0.635	107	49	2.355	49	2.355	1
5sga	20	0.791	67	30	20.488	29	1.263	2
6tmn	17	1.009	81	27	2.399	27	2.399	1

Bei dem Vergleich der Domänen durch den *Geometric Hashing*-Algorithmus entstehen viele Strukturen mit einer sehr ähnlichen Ligandenorientierung. Ist die Anzahl der zur Bestimmung der Struktur herangezogenen Domänenpaare groß, kann diese durch eine Vielzahl von Tripeln topographischer Domänenpaare gebildet werden. Da alle diese Tripel, die eine ausreichende Komplementarität aufweisen, zur Ermittlung der Komplexstrukturen herangezogen werden, resultiert daraus die Vielzahl ähnlicher Ligandenorientierungen. Durch den Clusteralgorithmus können diese Orientierungen dann zu einer mittleren Struktur zusammengefaßt werden. Als maximaler Abstand wurde dabei ein Wert von 3.0 Å verwendet. Die Ergebnisse des Algorithmus sowie die benötigte Rechenzeit sind in Tabelle 5.9 zusammengestellt. Es zeigt sich, daß bei gleichbleibender Qualität der *Docking*-Ergebnisse die Anzahl der Komplexstrukturen stark verringert werden kann. Dabei liegen vor allem um die kristallographisch bestimmte Komplexstruktur sehr viele Strukturen, die zu einer gemittelten zusammengefaßt werden können.

Alle so entstandenen Cluster von Komplexstrukturen wurden anschließend einer Energieminimierung mit der *Downhill Simplex*-Methode unterzogen. Die Ergebnisse sind in Tabelle 5.10 zusammengestellt. Die Berechnungen ergaben für alle untersuchten Komplexe mit nur zwei Ausnahmen eine der kristallographischen Struktur sehr ähnliche Strukturvorhersage als energieärmsten Komplex. Diese Struktur hebt sich energetisch durch einen Abstand von mehreren zehn Energieeinheiten deutlich von den weiteren Komplexvorhersagen ab.

Tabelle 5.9: Ergebnisse des Clusteralgorithmus für Enzym-Ligand-Komplexe mit niedermolekularen Inhibitoren (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	Anzahl der Cluster	<i>rms</i> -Wert (Å) des besten Clusters	Anzahl Strukturen im besten Cluster	Rechenzeit (s)
<b>1hne</b>	314	1.189	7	0.36
<b>1tlp</b>	668	1.308	28	1.34
<b>2er9</b>	9233	1.287	22	42.32
<b>2est</b>	129	2.323	3	0.08
<b>3apr</b>	7156	1.289	104	37.67
<b>4dfr</b>	311	0.802	13	0.79
<b>4hvp</b>	4328	1.632	70	114.98
<b>5apr</b>	5190	2.154	21	27.28
<b>5sga</b>	314	0.438	23	0.53
<b>6tmn</b>	352	2.085	8	0.28

In den zwei Ausnahmefällen (Komplexe 1hne und 6tmn) lag die zur kristallographisch bestimmten Kristallstruktur ähnlichste Vorhersage unter den zehn Bindungsmodi mit den geringsten Energien. Im Gegensatz zu den anderen Komplexen haben hier die energieärmsten Strukturen eine sehr ähnliche Energie, so daß sich nicht eine einzelne Struktur als Vorhersage des Komplexes auszeichnet. So ist der Abstand zwischen der energieärmsten Struktur (-78 a.u.) des Komplexes 1hne nur 13 a.u. energetisch günstiger als die Struktur mit dem geringsten *rms*-Wert (-64 a.u.). Im Fall des Komplexes 6tmn sind es sogar nur 4 a.u. (energieärmste Struktur: -86 a.u.; Struktur mit geringstem *rms*-Wert: -82 a.u.).

Tabelle 5.10: Ergebnisse der rigiden Energieminimierung mit dem *Downhill Simplex*-Algorithmus für Enzym-Ligand-Komplexe mit niedermolekularen Inhibitoren (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	<i>rms</i> -Wert (Å) der besten Struktur	Rang der besten Struktur	Energie der besten Struktur	Rechenzeit (min:s)
<b>1hne</b>	1.582	10	-64	0:34
<b>1tlp</b>	0.549	1	-101	1:22
<b>2er9</b>	0.140	1	-168	26:52
<b>2est</b>	0.592	1	-96	0:22
<b>3apr</b>	0.112	1	-170	22:02
<b>4dfr</b>	0.354	1	-114	0:30
<b>4hvp</b>	0.356	1	-153	9:32
<b>5apr</b>	0.185	1	-131	12:28
<b>5sga</b>	0.374	1	-124	0:31
<b>6tmn</b>	0.399	5	-81	0:49

Mit dem *Downhill Simplex*-Algorithmus wurden einige der vorhergesagten Komplexstrukturen in das gleiche lokale Minimum optimiert. Deswegen konnte durch den Clusteralgorithmus die Anzahl der Strukturen noch einmal verringert werden. Dabei sollten aber nur Strukturen zusammengefaßt werden, die eine sehr ähnliche Orientierung aufwiesen.



Deshalb wurde der maximale Abstand für eine Komplexstruktur zur mittleren Struktur des Clusters auf einen *rms*-Wert von 2.0 Å festgelegt. Die so erhaltenen Endergebnisse für die Enzym-Ligand-Komplexe mit dem starren *Docking*-Algorithmus sind in Tabelle 5.11 zusammengefaßt.

Tabelle 5.11: Endergebnisse des starren *Dockings* der Enzym-Ligand-Komplexe mit niedermolekularen Inhibitoren (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	Anzahl der Cluster	<i>rms</i> -Wert (Å)	Rang	Rechenzeit Clustering (s)	Gesamt-Rechenzeit (min:s)
<b>1hne</b>	278	1.582	6	0.12	0:50
<b>1tlp</b>	552	0.575	1	0.25	2:23
<b>2er9</b>	7323	0.226	1	6.04	37:50
<b>2est</b>	121	0.592	1	0.09	0:40
<b>3apr</b>	5439	0.172	1	4.30	28:01
<b>4dfr</b>	236	0.371	1	0.14	0:50
<b>4hvp</b>	2589	0.412	1	3.22	15:01
<b>5apr</b>	4011	0.277	1	3.25	17:07
<b>5sga</b>	271	0.366	1	0.19	0:49
<b>6tmn</b>	315	0.401	3	0.12	1:23

### 5.3.2 Komplex der Dihydrofolat-Reduktase mit Methotrexat (4dfr)

Eine der Vorstufen in der Synthese der Nucleotidbase Thymin ist das Desoxythymidilat. Zu dessen Synthese wird der Cofaktor Tetrahydrofolat in Dihydrofolat umgesetzt, welches durch das Enzym Dihydrofolat-Reduktase wieder zu Tetrahydrofolat regeneriert wird. Durch den an Dihydrofolat-Reduktase bindenden Inhibitor Methotrexat (Abb. 5.5) wird die Regenerierung des Tetrahydrofolat blockiert. Dadurch ist auch die Synthese von Thymin nicht mehr im für eine Zellteilung nötigen Maße möglich. Diese Blockierung führt so zu einer toxischen Wirkung, der vor allem Zellen mit hoher Teilungsrate, wie z.B. Krebszellen, die aktive DNA synthetisieren, stark ausgesetzt sind. Durch das spezifisch bindende Pteridin-Ringsystem, den fast unverzweigten Aufbau des Liganden und der charakteristischen Form des aktiven Zentrums eignet sich das System aus Dihydrofolat-Reduktase und Methotrexat (4dfr) für eine Überprüfung der Leistungsfähigkeit von *Docking*-Programmen und war deswegen ein Schwerpunkt dieser Untersuchung.

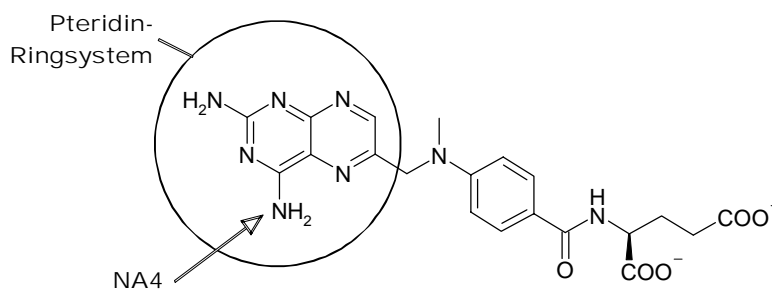


Abb. 5.5: Strukturformel von Methotrexat

Die kristallographische Komplexstruktur der Dihydrofolat-Reduktase und Methotrexat ist in Abbildung 5.6 dargestellt. Der Rezeptor wird durch die Gesamtfläche der als komplementär erkannten topographischen Domänen repräsentiert. Es zeigt sich, daß fast die gesamte Oberfläche des aktiven Zentrums als komplementär zum Liganden erkannt wird. Die verwendete Domäneneinteilung und der anschließende Vergleichsalgorithmus kann somit für die Beschreibung der Komplementarität herangezogen werden.

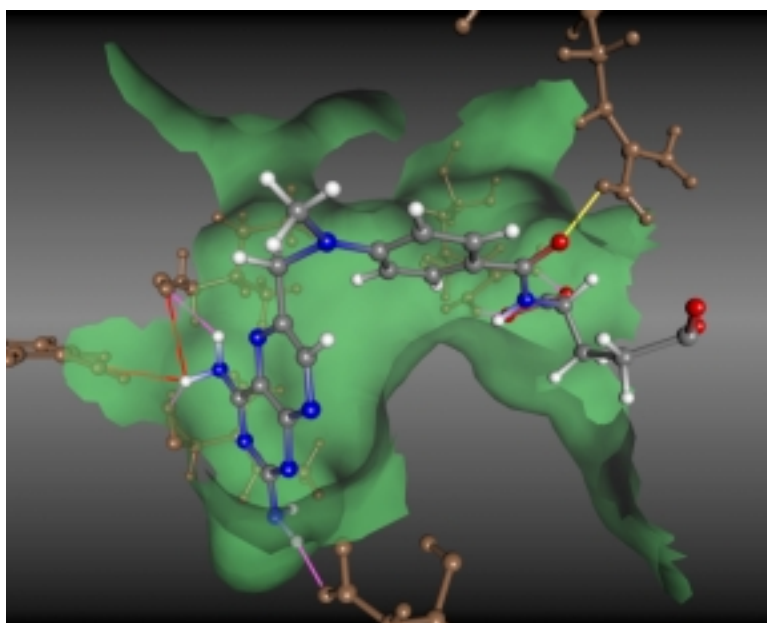


Abb. 5.6 Darstellung des Bindungsmodus des Komplexes aus Dihydrofolat-Reduktase und Methotrexat

Neben der durch den *Docking*-Algorithmus ermittelten komplementären Oberfläche des Rezeptors sind die Wasserstoffbrückenbindungen zwischen dem Liganden und dem Rezeptor gezeigt. Die durch den *Docking*-Algorithmus zusätzlich gefundenen Wasserstoffbrücken sind in rot dargestellt, die nur durch die Abstands- und Winkelkriterien [169] ermittelte in gelb. Wasserstoffbrücken, die von beiden Algorithmen erkannt werden, sind als magenta-farbene Verbindungslinien dargestellt.

Abbildung 5.6 zeigt ebenfalls die zwischen dem Liganden und dem Rezeptor gebildeten Wasserstoffbrückenbindungen. Zur Berechnung der gebildeten Wasserstoffbrückenbindungen

wurden die Abstands- und Winkelkriterien von Pirard et al. [169] verwendet. Der maximale Abstand zwischen dem Akzeptor und dem Donor wurde auf 3.2 Å und der Abstand zwischen dem Akzeptor und dem beteiligten Wasserstoff auf 3.0 Å festgelegt. Der Winkel am Wasserstoffatom muß größer als 90° sein. Durch den hier vorgestellten *Docking*-Algorithmus konnten fünf der sechs Wasserstoffbrücken vorhergesagt werden. Daneben wurden zwei weitere mögliche Wasserstoffbrückenbindungen vorgeschlagen. Als Donor wirkt der NA4-Stickstoff des Pteridin-Ringsystems, der eine Wasserstoffbrücke mit dem Amid-Sauerstoffatom der Aminosäure ILE94 bzw. dem OH-Sauerstoff von TYR100 ausbildet. Durch die oben beschriebenen Abstands- und Winkelkriterien wird eine Wasserstoffbrücke zwischen dem NA4-Stickstoff und dem Amid-Sauerstoff der Aminosäure ILE5 vorhergesagt. Da diese eine bessere geometrische Ausrichtung als die beiden anderen Möglichkeiten aufweist, ist sie als bevorzugte Wasserstoffbrücke anzusehen. Die im *Docking*-Algorithmus verwendete Methode beruht auf einem reinen Abstandskriterium zwischen dem Akzeptor und dem Wasserstoffatom und kann nicht zur energetischen Unterscheidung der drei Wasserstoffbrücken eingesetzt werden. Mit diesen Einschränkungen kann die Erkennung möglicher Wasserstoffbrückenbindungen durch den *Docking*-Algorithmus als geeignet angesehen werden.

### 5.3.3 Beschreibung weiterer Komplexstrukturen

Bei der Energieoptimierung mit dem *Downhill Simplex*-Algorithmus konnte für fast alle Komplexe die kristallographische Struktur als energieärmste Form ermittelt werden. Bei der Untersuchung von zwei Komplexen ergaben sich aber Strukturen, die trotz eines großen *rms*-Abstandes zur kristallographischen Komplexstruktur durch den Algorithmus als energetisch günstige Struktur erkannt wurden. Im Komplex aus Thermolysin und Cbz-Gly<sup>P</sup>-(O)-Leu-Leu (6tmn) konnte das aktive Zentrum des Enzyms zwar richtig erkannt werden, der Ligand wurde aber in der niederenergetischen Struktur genau um 180° verdreht in das Enzym eingepaßt (Abb. 5.7). Die für die Energieminimierung verwendete Zielfunktion ist somit nicht für alle Komplexe geeignet. Bei einigen Komplexen ist es notwendig, neben der durch die Zielfunktion berücksichtigten sterischen Komplementarität auch die elektrostatische Komplementarität sowie die explizite Ausbildung von Wasserstoffbrückenbindungen zu untersuchen. Die verwendete Zielfunktion berücksichtigt Wasserstoffbrücken, indem sich die beteiligten Atome auf einen geringeren Abstand annähern können und diese Atomkontakte stärker zur Energiebilanz beitragen als unpolare Wechselwirkungen. Für die explizite Behandlung der Wasserstoffbrücken muß aber auch die relative Lage von Akzeptor und Donor sowie dem Wasserstoffatom zueinander berücksichtigt werden, da die Energie einer Wasserstoffbrücke stark von den Bindungslängen und -winkeln abhängt.

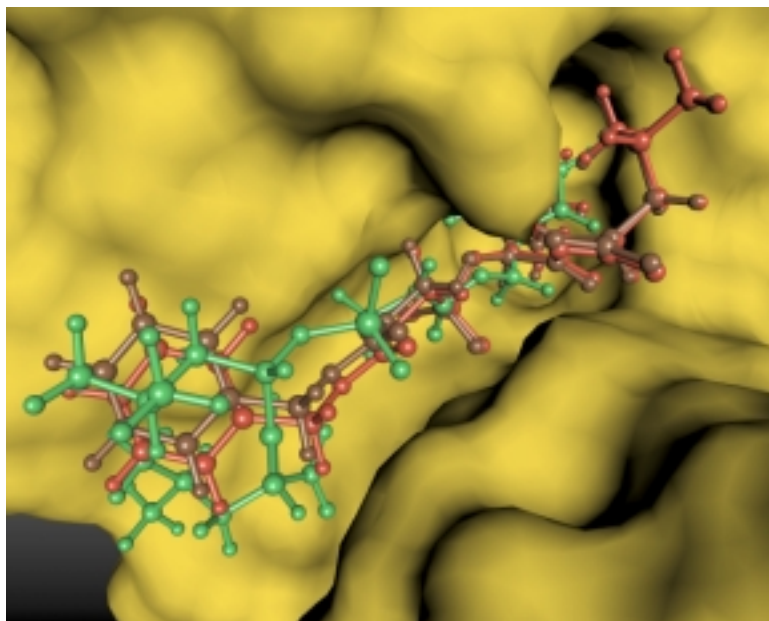


Abb. 5.7: Bindungsmodi des Komplexes aus Thermolysin und Cbz-Gly<sup>P</sup>-(O)-Leu-Leu  
Die kristallographisch bestimmte Struktur ist in rot gezeigt, die energieärmste in grün und die Struktur mit dem geringsten *rms*-Wert in braun. Es ist deutlich zu erkennen, daß der energieärmste Vorschlag genau um 180° um das aktive Zentrum gedreht ist.

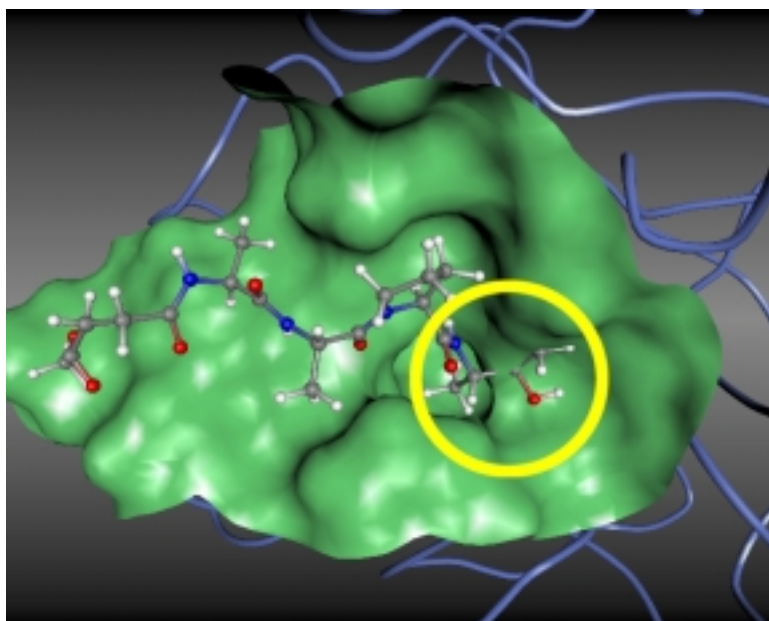


Abb. 5.8: Struktur des Komplexes der humanen Leukocyten-Elastase mit Methoxysuccinyl-Ala-Ala-Pro-Ala-Chromomethyl-Ketone  
Die Bindungsregion des Rezeptors ist durch die molekulare Oberfläche und der Ligand als *Ball-and-Stick*-Modell dargestellt. Der Ligand durchdringt die Oberfläche des Rezeptors (gelber Kreis). Dies wird durch den Algorithmus als Volumenüberlappung behandelt.

Anders sind die Verhältnisse beim Komplex der humanen Leukocyten-Elastase mit Methoxysuccinyl-Ala-Ala-Pro-Ala-Chloromethyl-Ketone. Der Inhibitor bildet in der Komplexstruktur eine kovalente Bindung zum Enzym aus. Da der vorgestellte Algorithmus davon ausgeht, daß der Ligand nur durch nichtbindende Wechselwirkungen in seiner Position fixiert wird, wird diese kovalente Bindung als starke Volumenüberlappung des Liganden mit dem Enzym behandelt (Abb. 5.8). Um den Einfluß der Volumenüberlappung auf den Algorithmus zu vermeiden, mußte der Proberadius zur Generierung der molekularen Oberfläche von 1.4 auf 1.35 Å verringert werden. Die relativ große *rms*-Abweichung von 1.582 Å und die im Vergleich zu anderen Komplexen wenig negative Energie von -64 a.u. sind auch auf die Vernachlässigung der Beiträge der kovalenten Bindung zurückzuführen.

### 5.3.4 Vorhersage für Enzym-Ligand-Komplexe mit inhibierenden Proteinen, Protein-Dimere und Antikörper-Antigen-Komplexe

Auch für die anderen Komplextypen wurde der *Docking*-Algorithmus wie im vorhergehenden Kapitel beschrieben durchgeführt. Dabei konnten die sehr guten Ergebnisse der niedermolekularen Liganden nicht reproduziert werden. Die Ergebnisse für die einzelnen Teilschritte des Algorithmus sind in Anhang 5 und zusammenfassend in Tabelle 5.12 aufgeführt. Für jeden Komplex ergab sich mit Hilfe des *Geometric Hashing*-Algorithmus aber mindestens eine Struktur, die einen *rms*-Wert von unter 5.0 Å zur Kristallstruktur aufweist. Dieser größere Abstand zur kristallographisch bestimmten Struktur ergibt sich einerseits aus den größeren Oberflächendimensionen und der daraus resultierenden größeren Unschärfe bei der Lage der komplementären Oberflächenreferenzpunkte. Andererseits resultiert bei diesen größeren Liganden aus einer kleinen Rotation ein größerer *rms*-Wert als bei kleinen Liganden.

Im ersten Teilschritt des Algorithmus ergibt sich für jeden Komplex eine sehr große Anzahl von möglichen Komplexstrukturen. Diese entstehen durch die Größe der beiden Komplexpartner und den daraus resultierenden Möglichkeiten der relativen Orientierung des Liganden zum Rezeptor. Zur effektiven Weiterverarbeitung mußte die Anzahl zuerst durch den in Kapitel 4.2.4 beschriebenen Test auf sterische Überlappung reduziert werden. Dazu wurde für alle vorhergesagten Komplexstrukturen die Volumenüberlappung im Verhältnis zur Oberflächenüberlappung berechnet und alle Strukturen gestrichen, die eine so definierte sterische Überlappung von  $B_{steric} \geq 0.04$  aufweisen. Dadurch konnte die Anzahl der Vorhersagen bei gleichbleibender Qualität der Ergebnisse um 60-80 % verringert werden. Im anschließenden Clustering wurden die größeren *rms*-Werte zwischen den Orientierungen berücksichtigt, indem der maximale Abstand zur mittleren, den Cluster repräsentierenden Struktur auf 5.0 Å gesetzt wurde.

Durch die abschließende Energieminimierung mit dem *Downhill Simplex*-Algorithmus konnte für jeden Komplex wieder eine Vorhersage gemacht werden, die der kristallographischen Struktur sehr ähnlich ist. Für alle bis auf einen Komplex entspricht diese Struktur auch der Vorhersage mit der geringsten Energie. Im Fall des Komplexes 4cpa ist die ähnlichste Struktur mit einem Energieunterschied von unter einer Energieeinheit die energetisch zweitniedrigste Struktur.

Tabelle 5.12: Endergebnisse des rigiden *Dockings* der Enzym-Ligand-Komplexe mit inhibierenden Proteinen, Protein-Dimere und Antikörper-Antigen-Komplexe (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	Anzahl der Strukturen	<i>rms</i> -Wert (Å)	Rang	Gesamt-Rechenzeit (h:min)
1acb	4354	1.190	1	1:51
1cbw	4926	0.523	1	2:08
1cho	2778	0.538	1	0:50
1cse	6332	0.834	1	2:42
1ppf	2617	1.176	1	0:59
1tec	5927	0.634	1	2:30
1tgs	3532	0.678	1	1:15
1tpa	3112	0.650	1	1:05
2cpk	1188	0.526	1	0:21
2kai	2475	0.600	1	1:02
2ptc	3494	0.679	1	1:19
2sec	4624	0.736	1	1:57
2sic	9472	1.388	1	5:40
2sni	4427	0.236	1	2:01
2tgp	3942	0.705	1	1:25
4cpa	3691	2.535	2	1:07
4sgb	2015	0.453	1	0:33
4tpi	3133	0.455	1	1:08
1aar	3372	0.985	1	0:48
2ccy	8670	0.360	1	4:27
2utg	3327	0.697	1	0:56
1fdl	12887	0.704	1	18:39
3hfl	17258	0.876	1	23:08
3hfm	22859	0.471	1	30:04
2igf	4512	0.354	1	0:30

Durch die Erweiterung des *Docking*-Algorithmus auf die verschiedenen Komplexarten konnte gezeigt werden, daß der Ansatz nicht von der bestimmten Form der Bindungsregionen abhängt. Da das aktive Zentrum der Enzyme meist in einer tiefen konkaven Tasche der molekularen Oberfläche liegt, könnte durch die Bevorzugung dieser Oberflächenbereiche eine starke Verringerung der vorgeschlagenen Komplexstrukturen erzielt werden. Im Gegensatz dazu sind die Bindungsregionen in Protein-Dimeren und Antikörper-Antigen-Komplexen relativ große und flache Oberflächenbereiche. Nur durch die Gleichbehandlung sämtlicher Oberflächendomänen kann gewährleistet werden, daß der Algorithmus für alle Komplexarten zu gleich guten Ergebnissen führt. Dies konnte für den hier beschriebenen Algorithmus

bestätigt werden. Sind über einen bestimmten Komplex weitere biochemische Informationen erhältlich, könnten diese nach dem *Geometric Hashing*-Algorithmus zur Verringerung der Anzahl der weiter zu untersuchenden Komplexstrukturen herangezogen werden. Bei den Enzym-Ligand-Komplexen würden z.B. nur solche Strukturen weiter behandelt, bei denen der Ligand in einer großen Oberflächentasche oder in der Nähe einer als katalytisch wirkenden Aminosäure des Enzyms liegt. Dadurch könnte die für die Energieoptimierung benötigte Zeit noch einmal stark verringert werden.

### 5.3.5 Bewertung des vorgeschlagenen *Docking*-Algorithmus

Die oben beschriebenen Ergebnisse zeigen, daß der hier vorgestellte *Docking*-Algorithmus zur schnellen Vorhersage von Komplexstrukturen geeignet ist. Dazu müssen lediglich die Strukturen der Komplexpartner bekannt sein. Weitere Informationen, wie z.B. die Lage des aktiven Zentrums, werden dagegen nicht benötigt. Dadurch unterscheidet er sich von den meisten in der Literatur vorgestellten Algorithmen (z.B. [12,170-172]). Für einen bestimmten Rezeptor reicht eine einmalige Anwendung aus, um das aktive Zentrum zu finden. Ist somit die Lage der Bindungsregion bekannt, können die Komplexstrukturen einer großen Anzahl möglicher Liganden (z.B. aus einer Strukturdatenbank) mit diesem Rezeptor mit Hilfe anderer Algorithmen ermittelt werden, die dann nur noch das aktive Zentrum des Rezeptors berücksichtigen.

Der Hauptvorteil des hier vorgestellten Algorithmus liegt in der Möglichkeit, die Bindungsregionen der beiden Komplexpartner sehr schnell zu definieren. Dies konnte an mehreren Beispielen mit bekannten Komplexstrukturen gezeigt werden. Schon der Vergleich der in Domänen eingeteilten molekularen Oberflächen der beiden Moleküle erlaubt die Vorhersage von Ligandenorientierungen, die eine ähnliche Lage wie in der Kristallstruktur aufweisen. Dabei zeigt sich vor allem bei niedermolekularen Liganden, daß diese Strukturvorhersagen durch eine große Anzahl von komplementären Domänenpaaren gebildet werden. Die Beschreibung der molekularen Oberflächen und der topographischen und physikochemischen Eigenschaften durch die Oberflächendomänen kann somit als eine weitgehendst eindeutige Beschreibung der Bindungsregionen verwendet werden, die empfindlich auf die beim *Docking* erzielbare Übereinstimmung anspricht. Da die Domäneneinteilung von der zur Segmentierung herangezogenen topographischen bzw. physikochemischen Eigenschaft unabhängig ist, kann diese auf weitere Eigenschaften ausgedehnt werden. Es ist dadurch ebenfalls möglich, Verbesserungen in den Berechnungsmethoden der einzelnen Eigenschaften direkt in den Algorithmus zu integrieren. Dies könnte vor allem für die Lipophilie in naher Zukunft relevant werden, da an einer Verbesserung der Beschreibung dieser Eigenschaft zur Zeit intensiv geforscht wird [173].

Das vorgeschlagene Verfahren bewertet in einem zweiten Schritt die so generierten Strukturen in Bezug auf die eindeutige Charakterisierung des aktiven Zentrums. Diese Bewertung läßt sich nicht direkt mit der Anzahl komplementärer Domänenpaare erreichen, da für Komplexe mit großen Liganden eine große Anzahl von vorgeschlagenen Komplexstrukturen die gleiche oder sogar eine noch größere Anzahl von komplementären Domänen aufweist. Deswegen werden durch eine sehr einfache, schnell auswertbare Energiefunktion die relativen Komplexbindungsenergien berechnet, um daraus eine energetische Rangfolge der Strukturen zu bestimmen. Durch die unscharfe Definition der Bindungsregionen weisen die Komplexstrukturen meistens noch eine sterische Überlappung auf, die vor der Energiebestimmung durch eine Optimierung der Struktur eliminiert werden muß. Durch die Optimierung mit dem *Downhill Simplex*-Algorithmus und die anschließende Energiebestimmung konnten für fast alle der in dieser Arbeit untersuchten Komplexe die kristallographische Struktur als beste Vorhersage ermittelt werden. Da diese Energieminimierung aber rechentechnisch aufwendig ist, sollen Strukturen, die nicht in der Nähe eines lokalen Minimums liegen, nicht weiter berücksichtigt werden. Durch den für große Liganden durchgeführten Test auf starke sterische Überlappung konnte die Anzahl der Komplexstrukturen und dadurch die Rechenzeit für die anschließende Optimierung stark reduziert werden. In diesem Schritt liegt auch das größte Potential zur Verbesserung des vorgeschlagenen *Docking*-Algorithmus. Durch die direkte energetische Bewertung der aus dem *Geometric Hashing*-Algorithmus resultierenden Strukturen könnte die Anzahl der weiter zu untersuchenden Strukturen weiter stark verringert werden. Dazu müßte aber die unscharfe Plazierung des Liganden in dieser Energiebewertung berücksichtigt werden.

#### 5.4 Flexibles *Docking* von Methotrexat in Dihydrofolat-Reduktase

Im letzten Teil der Arbeit wurde getestet, ob der *Docking*-Algorithmus prinzipiell in der Lage ist, Komplexe mit unbekannter Komplexstruktur vorherzusagen. Dazu wurde an dem bereits oben beschriebenen Komplex aus Dihydrofolat-Reduktase und Methotrexat (Kapitel 5.3.2) ein *Docking*-Experiment unter Berücksichtigung der Flexibilität des Liganden durchgeführt. Dabei wurde nur die kristallographisch bestimmte Struktur des Rezeptors verwendet. Für den Liganden wurden der Kristallstruktur keine Informationen über dessen Konformation entnommen.

Im ersten Schritt wurden mehrere niederenergetische Konformationen des Liganden berechnet. Dies erfolgte mit dem im Programmpaket *SYBYL* enthaltenen Modul *GA Conformational Search* [174]. Dabei werden ein genetischer Algorithmus und das Tripos Kraftfeld [175] als *Fitness*-Funktion zum Auffinden dieser niederenergetischen Konformationen verwendet. Alle auch bei der flexiblen Optimierung der *Docking*-Strukturen



als flexibel angesehenen Einfachbindungen wurden in die Bestimmung niederenergetischer Konformationen einbezogen, Bindungslängen und Bindungswinkel wurden konstant gehalten. Es wurden 200 Individuen über 5000 Generationen optimiert. Die besten zwanzig Konformationen wurden als Ausgangsstrukturen für den *Docking*-Algorithmus verwendet und sind in Anhang 6 dargestellt. Um eine möglichst komplette Beschreibung des Konformationsraums zu erhalten, wird durch den Algorithmus überprüft, daß sich zwei Konformationen nicht zu ähnlich werden.

Für alle diese so bestimmten Konformationen des Liganden wurde der oben beschriebene starre *Docking*-Algorithmus unabhängig von den anderen Konformationen durchgeführt. Dazu wurden zuerst die molekularen Oberflächen sowie deren topographische Eigenschaften und die physikochemischen Eigenschaften am Ort der molekularen Oberfläche für jede Konformation des Liganden sowie der kristallographisch bestimmten Komplexstruktur des Rezeptors berechnet. Die Oberflächen wurden in Domänen eingeteilt und diese mit dem *Geometric Hashing*-Algorithmus verglichen. Die daraus resultierenden Komplexstrukturen wurden in Cluster eingeteilt und mit dem *Downhill Simplex*-Verfahren unter Verwendung starrer Molekülstrukturen energieminiert. Die Ergebnisse der Einzelschritte für jede Ligandenkonformation sind ebenfalls in Anhang 6 aufgeführt. Durch das starre *Docking* konnte eine Vorhersage der Komplexstruktur mit einer Energie von  $-83$  a.u. und einem *rms*-Wert von  $1.203 \text{ \AA}$  ermittelt werden, die schon als sehr gute Lösung der Komplexstruktur angesehen werden kann. Der Energiewert dieser besten Struktur liegt aber noch deutlich über dem mit der kristallographisch bestimmten Komplexstruktur berechneten Wert von  $-114$  a.u. Nach dem starren *Docking* aller Ligandenkonformationen konnten insgesamt 15 Komplexstrukturen bestimmt werden, die eine niedrigere Energie als diese ähnlichste Struktur aufweisen.

Für eine weitere Verbesserung der Komplexstruktur wurde die zur kristallographischen Struktur ähnlichste Vorhersage unter Berücksichtigung der Flexibilität des Liganden mit dem genetischen Algorithmus weiter optimiert. Da im aktiven Zentrum zwei Wassermoleküle (HOH 403 und 405 in *pdb*-Notation) relativ stark an das Enzym binden und diese in der Kristallstruktur Wasserstoffbrücken zwischen dem Pteridin-Ring des Methotrexat und dem Enzym vermitteln, wurde für die Struktur des Rezeptors die Enzymstruktur mit allen kristallographisch bestimmten Positionen von Wassermolekülen verwendet. Diese Optimierung erfolgte auf einer SGI ONIX mit R10000 Prozessor und benötigte eine Rechenzeit von 1:40 Stunden. Die beste Struktur konnte auf einen *rms*-Wert von  $1.056 \text{ \AA}$  (Abb. 5.9) verbessert werden. Für die Energie ergab sich ein Wert von  $-122$  a.u.

Mit Hilfe der an diesem Beispiel erhaltenen Ergebnisse konnte gezeigt werden, daß sich der Algorithmus auch für die Vorhersage von Komplexen eignet, in denen nur die Struktur der

ungebundenen Komplexpartner bekannt ist. So konnte mit einer sehr kleinen Anzahl von Ausgangskonformationen des Liganden die kristallographisch bestimmte Komplexstruktur reproduziert werden. Auch für die Ausgangskonformationen, die sich stark von der kristallographisch bestimmten Struktur unterscheiden, konnten dabei Teillösungen der Komplexstruktur ermittelt werden. Für das flexible *Docking* gelten aber alle oben erwähnten Einschränkungen gleichermaßen (siehe Kapitel 5.3.2 bis 5.3.5). Die Energieminimierung mit dem *Downhill Simplex*-Algorithmus und dem anschließenden genetischen Algorithmus führte zu einer guten Übereinstimmung mit der kristallographischen Komplexstruktur.

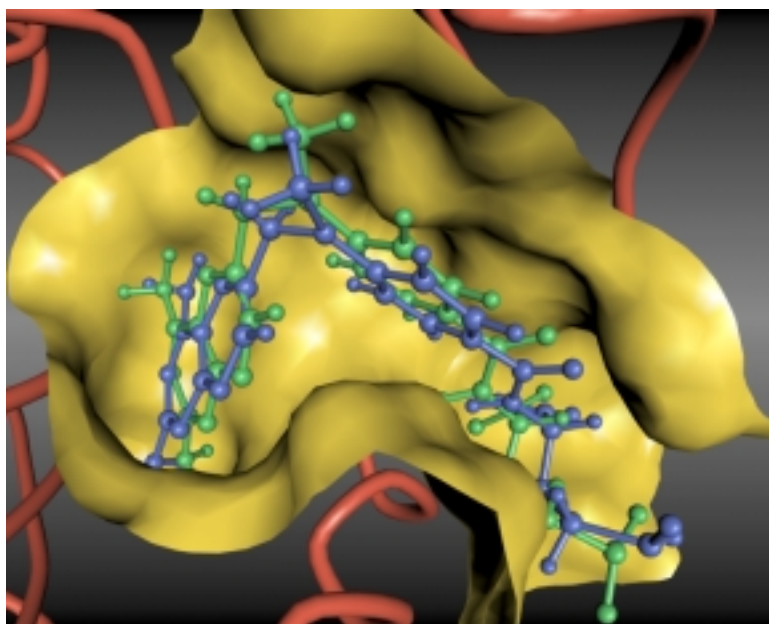


Abb. 5.9: Vergleich der kristallographisch bestimmten und der mit Hilfe des flexiblen *Docking*-Algorithmus bestimmten Komplexstruktur von Dihydrofolat-Reduktase mit Methotrexat. Die kristallographisch bestimmte Ligandenorientierung ist als blaues und die durch den Algorithmus bestimmte als grünes *Ball-and-Stick*-Modell dargestellt. Das Enzym wird durch die molekulare Oberfläche des aktiven Zentrums und durch eine *Ribbon*-Darstellung des Aminosäuren-Rückgrats repräsentiert.

Durch die Berücksichtigung weiterer Kriterien sollte eine weitere Verbesserung der Orientierung, vor allem eine Verbesserung der Lage des Pteridin-Ringsystems (siehe Abbildung 5.5) durch eine Optimierung der Wasserstoffbrückengeometrien, möglich sein. Zusätzlich sollten sich durch die Optimierung der Zielfunktion die Bindungsregionen mit dem *Downhill Simplex*-Algorithmus eindeutig charakterisieren lassen. Ist dies nicht der Fall, muß der rechenintensive genetische Algorithmus auf mehrere Komplexstrukturen angewendet werden, wodurch sich die Effizienz des Algorithmus stark verringert. Deswegen sollte die Zielfunktion in der Lage sein, die Unterschiede in den Atompositionen der Ausgangskonformationen und der kristallographischen Konformation zu berücksichtigen. Dies könnte durch die Einführung einer gewissen Unschärfe der Atompositionen in die

Zielfunktion erfolgen. Bei der Implementierung des genetischen Algorithmus sowie der *Downhill Simplex*-Methode wurde darauf geachtet, daß die verwendete Zielfunktion ohne großen programmiertechnischen Aufwand durch eine andere ersetzt werden kann. Dadurch können Verbesserungen an der Zielfunktion ohne Veränderung des Programmablaufes vorgenommen werden.

## Kapitel 6

### Zusammenfassung

In der hier vorgestellten Arbeit wird ein *Docking*-Algorithmus entwickelt, der ausgehend von den dreidimensionalen Strukturdaten biochemischer Komplexpartner das aktive Zentrum eines Rezeptors identifiziert und Vorschläge für mögliche Bindungsmodi eines Liganden in diesem aktive Zentrum macht. Dazu wurde versucht, die menschliche Fähigkeit der Formerkennung und Problemoptimierung in mathematischen Algorithmen abzubilden. Die Grundlage dazu bildet die *Fuzzy Set Theory*, die eine Beschreibung verbaler Begriffe in mathematisch zugänglicher Form, den linguistischen Variablen, ermöglicht.

Der Algorithmus beruht auf dem Modell einer molekularen Oberfläche. Jeder Punkt dieser Oberfläche wird durch die topographischen Eigenschaften, das elektrostatische Potential, die Lipophilie und die Fähigkeit, Wasserstoffbrückenbindungen auszubilden, charakterisiert. Den Eigenschaften werden linguistische Variablen zugeordnet, mit deren Hilfe die Ähnlichkeit zwischen Oberflächenpunkten berechnet werden kann. Diese Ähnlichkeit wird dazu benutzt, die Oberfläche in größere Bereiche mit ähnlichen Eigenschaften einzuteilen. Die Grenzen der dabei entstehenden, als Domänen bezeichneten Bereiche werden über eine Zugehörigkeitsfunktion definiert, die von einem Wert von 1 für vollständige Zugehörigkeit bis zu einem Wert von 0 für Punkte, die nicht zur Domäne gehören, läuft. Dadurch ist es möglich, daß Oberflächenpunkte zu mehreren Domänen gehören und daß sich die Domänen überschneiden. Außerdem kann dadurch die Größe der Domänen aufeinander angepaßt werden.

Mit Hilfe des *Geometric Hashing*-Algorithmus [26,27] werden mögliche Komplexstrukturen bestimmt. Dazu werden die Oberflächendomänen durch charakteristische Punkte und einen assoziierten Normalenvektor beschrieben und die dreidimensionale Lage dieser charakteristischen Punkte beider Moleküle miteinander verglichen. Die charakteristischen Punkte werden als Mittelpunkte der nach den topographischen Eigenschaften, dem elektrostatischen Potential und der Lipophilie eingeteilten Domänen sowie der Lage möglicher Wasserstoffbrückenbindungspartner auf den molekularen Oberflächen definiert. Die Komplementarität der Eigenschaften der Kombinationen einer Domäne des Liganden mit einer Domäne des Rezeptors ergibt sich durch ein *Fuzzy*-Kriterium, das über die zu jeder Eigenschaft zugeordneten linguistischen Variablen berechnet wird. Charakteristische Punkte der so bestimmten komplementären Domänen mit entsprechender dreidimensionaler Anordnung werden mit einem *Least Square Fit*-Algorithmus [113] aufeinander projiziert. Mit diesem Verfahren konnten so sehr schnell erste Vorhersagen für die Komplexstruktur gemacht werden.

Da durch den *Geometric Hashing*-Algorithmus eine große Auswahl von möglichen Komplexstrukturen vorgeschlagen werden, muß deren Anzahl vor der energetischen Bewertung der Vorhersagen verringert werden. Dazu werden zuerst Strukturen, die eine starke sterische Überlappung aufweisen, verworfen. Für die Berechnung der sterischen Überlappung wurde eine empirische Funktion entwickelt, welche die Anzahl der überlappenden Atome mit der Größe der Bindungsregion wichtet. Ähnliche Komplexstrukturen werden über einen Clusteralgorithmus zu einer mittleren Struktur zusammengefaßt. Dazu wird ein agglomeratives hierarchisches Verfahren eingesetzt, das als Abstandsfunktion zwischen zwei Clustern die *rms*-Abweichung der die Cluster repräsentierenden mittleren Ligandenorientierungen verwendet.

Für die energetische Bewertung der durch den Clusteralgorithmus erzeugten Komplexstrukturen wird die empirische Zielfunktion von Gehlhaar et al. [121] herangezogen. Die Strukturen werden mit dem *Downhill Simplex*-Verfahren [68] in das nächstgelegene Minimum überführt. Dabei wird nur die relative Lage der beiden Moleküle zueinander, nicht aber die Konformationen der einzelnen Moleküle optimiert. Strukturen, die bei der Energieoptimierung im selben Minimum enden, werden durch erneute Verwendung des Clusteralgorithmus zusammengefaßt und alle resultierenden Komplexstrukturen nach steigender Energie sortiert. Um die Flexibilität des Liganden zu berücksichtigen, wird eine energetisch günstige Struktur mit einem genetischen Algorithmus weiter optimiert. Dabei werden neben der relativen Orientierung auch die Torsionswinkel von frei rotierbaren Bindungen des Liganden optimiert. Als rotierbare Bindungen werden alle Einfachbindungen zwischen zwei  $sp^3$ -hybridisierten Atomen bzw. einem  $sp^3$ - und einem  $sp^2$ -hybridisierten Atom, die nicht Teil eines Ringsystems sind, angesehen.

Der Algorithmus wurde an 35 einer Datenbank entnommenen biochemischen Komplexen getestet. Dabei sind die Beispiele folgenden Arten von biochemischen Komplexen zuzuordnen: 10 Komplexe eines Enzyms mit einem niedermolekularen Inhibitor, 18 Komplexe eines Enzyms mit einem inhibierenden Protein, 3 Protein-Dimere und 4 Antikörper-Antigen-Komplexe. Zur Ermittlung der Komplexvorhersagen werden jeweils die Konformationen aus der kristallographisch bestimmten Komplexstruktur verwendet. Der *Geometric Hashing*-Algorithmus ist bereits in der Lage, Strukturen nahe an der kristallographisch bestimmten Komplexstruktur zu ermitteln. Dabei zeigt es sich, daß für die Komplexe mit niedermolekularen Inhibitoren die besten Ergebnisse erzielt werden. In diesen Komplexen zeichnen sich die kristallographisch bestimmten Bindungsregionen durch eine große Anzahl komplementärer Domänen aus. Deshalb kann unter den Komplexstrukturen mit den meisten Domänenpaaren mindestens eine Struktur mit einem *rms*-Wert unter 2.5 Å ermittelt werden. Die Ergebnisse können durch die anschließende Energieoptimierung mit dem *Downhill Simplex*-Algorithmus weiter verbessert werden. Bis auf drei Ausnahmen wird

für alle Komplexe die kristallographische Struktur als Vorhersage mit der geringsten Energie bestimmt. Die *rms*-Abweichung zur kristallographischen Struktur beträgt maximal 1.4 Å. Für die drei Ausnahmefälle konnte unter den zehn energieärmsten Strukturen eine Vorhersage mit einem *rms*-Wert unter 2.6 Å ermittelt werden. Die Gründe für die schlechteren Ergebnisse dieser drei Komplexe sind vor allem auf die Unzulänglichkeiten der empirischen Zielfunktion und den Umstand, daß keine kovalenten Bindungen zwischen dem Liganden und dem Rezeptor vorhergesagt werden können, zurückzuführen. Durch eine Erweiterung der Zielfunktion auf weitere Terme (Salzbrücken, elektrostatische Anziehung, entropische Effekte) kann deswegen eine Verbesserung des *Docking*-Algorithmus erzielt werden.

Um die Flexibilität des Liganden zu berücksichtigen, wird am Beispiel eines Enzym-Inhibitor-Komplexes (Dihydrofolat-Reduktase / Methotrexat, 4dfr) ein *Docking*-Experiment durchgeführt, ohne die Informationen über die Konformation des Liganden aus der Kristallstruktur zu verwenden. Dazu werden 20 niederenergetische Konformationen des Liganden erzeugt. Diese dienen neben der Konformation des Rezeptors aus der kristallographisch bestimmten Komplexstruktur als Eingabewerte für den beschriebenen *Docking*-Algorithmus. Bereits nach der starren Optimierung durch den *Downhill Simplex*-Algorithmus ergibt sich eine Struktur, die einen *rms*-Abstand von nur 1.2 Å aufweist. Der geringe *rms*-Wert zeigt, daß schon 20 Konformationen ausreichen, um mit dem hier vorgestellten Algorithmus das *Docking*-Problem zufriedenstellend zu lösen. Diese beste Struktur kann mit dem genetischen Algorithmus unter Berücksichtigung der Flexibilität des Liganden weiter optimiert werden. Auf diese Weise konnte die Abweichung zur kristallographischen Struktur weiter verringert werden.

Die Ergebnisse zeigen eindeutig, daß der Algorithmus zur Vorhersage biochemischer Komplexstrukturen geeignet ist. Durch die Erweiterung auf flexible oder teilflexible Konformationen wird auch die Vorhersage von Komplexen, bei denen nur die Strukturen der ungebundenen Komplexpartner bekannt sind, möglich. Die mit dem vorgestellten Verfahren mögliche Lokalisierung und Charakterisierung des aktiven Zentrums eines Rezeptormoleküls ist ein wesentlicher Beitrag zur computergestützten Untersuchung biochemischer Prozesse und damit ein weiterer Schritt auf dem Weg, die extrem Zeitintensität der konventionellen Medikamentenentwicklung zu reduzieren.

## Literatur

1. Böhm, H.-J.; Klebe, G.; and Kubinyi, H.; "Wirkstoffdesign"; Spektrum Akademischer Verlag GmbH; Heidelberg; **1996**.
2. Brickmann, J.; "Linguistic Variables in the Molecular Recognition Problem" in Fuzzy Logic in Chemistry; Academic Press; San Diego; **1996**; 1-47.
3. Zimmermann, H.-J.; "Fuzzy Set Theory and Its Applications"; Kluwer; Boston; **1991**.
4. Böhme, G.; "Fuzzy-Logik"; Springer Verlag; Berlin; **1993**.
5. Zadeh, L. A.; Fuzzy Sets; *Information and Control* (**1965**); 8; 338-353.
6. Stryer, L.; "Biochemie"; Spektrum der Wissenschaft-Verlagsgesellschaft; Heidelberg; **1990**.
7. Voet, D. and Voet, J. G.; "Biochemie"; Verlag Chemie; Weinheim; **1992**.
8. Fischer, E.; Einfluss der Configuration auf die Wirkung der Enzyme; *Chem.Ber.* (**1894**); 27; 2985-2993.
9. Sternberg, M. J. E.; Gabb, H. A.; and Jackson, R. M.; Predictive Docking of Protein-Protein and Protein-DNA Complexes; *Curr.Opin.Struct.Biol.* (**1998**); 8; 250-256.
10. Bamborough, P. and Cohen, F. E.; Modeling Protein-Ligand Complexes; *Curr.Opin.Struct.Biol.* ( **1996**); 6; 236-241.
11. Lengauer, Th. and Rarey, M.; Computational Methods for Biomolecular Docking; *Curr.Opin.Struct.Biol.* (**1996**); 6(3); 402-406.
12. Strynadka, N. C. J.; Eisenstein, M.; Katchalski-Katzir, E.; Shoichet, B. K.; Kuntz, I. D.; Abagyan, R.; Totrov, M.; Janin, J.; Cherfils, J.; Zimmerman, F.; Olson, A. J.; Duncan, B. S.; Rao, M.; Jackson, R. M.; Sternberg, M. J. E.; and James, M. N. G.; Molecular Docking Programs Successfully Predict the Binding of a  $\beta$ -Lactamase Inhibitory Protein to TEM-1  $\beta$ -Lactamase; *Nat.Struct.Biol.* (**1996**); 3(3); 233-239.
13. Kuntz, I. D.; Meng, E. C.; and Shoichet, B. K.; Structure-Based Molecular Design; *Acc.Chem.Res.* (**1994**); 27(5); 117-123.
14. Rarey, M.; Kramer, B.; and Lengauer, Th.; Multiple Automatic Base Selection: Protein-Ligand Docking Based on Incremental Construction Without Manual Intervention; *J.Comput.-Aided Mol.Des.* (**1997**); 11; 369-384.
15. Sobolev, V.; Wade, R. C.; Vriend, G.; and Edelman, M.; Molecular Docking Using Surface Complementarity; *Proteins* (**1996**); 25; 120-129.
16. Welch, W.; Ruppert, J.; and Jain, A. N.; Hammerhead: Fast, Fully Automated Docking of Flexible Ligands to Protein Binding Sites; *Chem.Biol.* (**1996**); 3(6); 449-463.
17. Desmet, J.; Wilson, I. A.; Joniau, M.; De Maeyer, M.; and Lasters, I.; Computation of the Binding of Fully Flexible Peptides to Protein With Flexible Side Chains; *FASEB J.* (**1997**); 11; 164-172.

18. Clark, K. P. and Ajay; Flexible Ligand Docking Without Parameter Adjustment Across Four Ligand-Receptor Complexes; *J.Comp.Chem.* (1995); 16(10); 1210-1226.
19. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; and Ferrin, Th. E.; A Geometric Approach to Macromolecule-Ligand Interactions; *J.Mol.Biol.* (1982); 161; 269-288.
20. Oshiro, C. M.; Kuntz, I. D.; and Dixon, J. S.; Flexible Ligand Docking Using a Genetic Algorithm; *J.Comput.-Aided Mol.Des.* (1995); 9; 113-130.
21. Gschwend, D. A. and Kuntz, I. D.; Orientational Sampling and Rigid-Body Minimization in Molecular Docking Revisited: On-the-Fly Optimization and Degeneracy Removal; *J.Comput.-Aided Mol.Des.* (1996); 10; 123-132.
22. Meng, E. C.; Shoichet, B. K.; and Kuntz, I. D.; Automated Docking With Grid-Based Energy Evaluation; *J.Comp.Chem.* (1992); 13(4); 505-524.
23. Connolly, M. L.; Shape Complementarity at the Hemoglobin Subunit Interface; *Biopolymers* (1986); 25; 1229-1247.
24. Hendrix, D. K. and Kuntz, I. D.; Surface Solid Angle-Based Site Points for Molecular Docking; *Proceedings of the Pacific Symposium on Biocomputing '96* (1997); 317-326.
25. Lin, S. L.; Nussinov, R.; Fischer, D.; and Wolfson, H. J.; Molecular Surface Representations by Sparse Critical Points; *Proteins* (1994); 18; 94-101.
26. Lamdan, Y.; Schwartz, J. T.; and Wolfson, H. J.; Affine Invariant Model-Based Object Recognition; *IEEE Transactions on Robotics and Automation* (1990); 6(5); 578-589.
27. Schwartz, J. T. and Sharir, M.; Identification of Partially Obscured Objects in Two and Three Dimensions by Matching Noisy Characteristic Curves; *Int.J.Robotics Res.* (1987); 6(2); 29-44.
28. Fischer, D.; Lin, S. L.; Wolfson, H. J.; and Nussinov, R.; A Geometry-Based Suite of Molecular Docking Processes; *J.Mol.Biol.* (1995); 248; 459-477.
29. Sandak, B.; Nussinov, R.; and Wolfson, H. J.; An Automated Computer Vision and Robotics-Based Technique for 3-D Flexible Biomolecular Docking and Matching; *CABIOS* (1995); 11(1); 87-99.
30. Sandak, B.; Wolfson, H. J.; and Nussinov, R.; Flexible Docking Allowing Induced Fit in Proteins: Insights From an Open to Closed Conformational Isomers; *Proteins* (1998); 32; 159-174.
31. Bacon, D. J. and Moulton, J.; Docking by Least-Squares Fitting of Molecular Surface Patterns; *J.Mol.Biol.* (1992); 225; 849-858.
32. Vakser, I. A.; Evaluation of GRAMM Low-Resolution Docking Methodology on the Hemagglutinin-Antibody Complex; *Proteins* (1997); Suppl 1; 226-230.



33. Ausiello, G.; Cesareni, G.; and Helmer-Citterich, M.; ESCHER: A New Docking Procedure Applied to the Reconstruction of Protein Tertiary Structure; *Proteins* (1997); 28; 556-567.
34. Helmer-Citterich, M. and Tramontano, A.; PUZZLE: A New Method for Automated Protein Docking Based on Surface Shape Complementarity; *J.Mol.Biol.* (1994); 235; 1021-1031.
35. Badel, A.; Mornon, J. P.; and Hazout, S.; Searching for Geometric Molecular Shape Complementarity Using Bidimensional Surface Profiles; *J.Mol.Graphics* (1992); 10; 205-211.
36. Katchalski-Katzir, E.; Shariv, I.; Einstein, M.; Friesem, A. A.; Aflalo, C.; and Vakser, I. A.; Molecular Surface Recognition: Determination of Geometric Fit Between Proteins and Their Ligands by Correlation Techniques; *Proc.Natl.Acad.Sci.USA* (1992); 89; 2195-2199.
37. Gabb, H. A.; Jackson, R. M.; and Sternberg, M. J. E.; Modelling Protein Docking Using Shape Complementarity, Electrostatics and Biochemical Information; *J.Mol.Biol.* (1997); 272; 106-120.
38. Meyer, M.; Wilson, P.; and Schomburg, D.; Hydrogen Bonding and Molecular Surface Shape Complementarity As a Basis for Protein Docking; *J.Mol.Biol.* (1996); 264; 199-210.
39. Caflisch, A.; Fischer, S.; and Karplus, M.; Docking by Monte Carlo Minimization With a Solvation Correction: Application to an FKBP-Substrate Complex; *J.Comp.Chem.* (1997); 18(6); 723-743.
40. Trosset, J.-Y. and Scheraga, H. A.; Reaching the Global Minimum in Docking Simulations: A Monte Carlo Minimization Approach Using Bezier Splines; *Proc.Natl.Acad.Sci.USA* (1998); 95; 8011-8015.
41. Hart, T. N.; Ness, S. R.; and Read, R. J.; Critical Evaluation of the Research Docking Program for the CASP2 Challenge; *Proteins* (1997); Suppl. 1; 205-209.
42. McMartin, C. and Bohacek, R. S.; QXP: Powerful, Rapid Computer Algorithms for Structure-Based Drug Design; *J.Comput.-Aided Mol.Des.* (1997); 11; 333-344.
43. Goodsell, D. S.; Morris, G. M.; and Olson, A. J.; Automated Docking of Flexible Ligands: Applications of AutoDock; *J.Mol.Recognit.* (1996); 9; 1-5.
44. Westhead, D. R.; Clark, D. E.; and Murray, Ch. W.; A Comparison of Heuristic Algorithms for Molecular Docking; *J.Comput.-Aided Mol.Des.* (1997); 11; 209-228.
45. Schaffer, L. and Verkhivker, G. M.; Predicting Structural Effects in HIV-1 Protease Mutant Complexes With Flexible Ligand Docking and Protein Side-Chain Optimization; *Proteins* (1998); 33(2); 295-310.

46. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; and Olson, A. J.; Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function; *J.Comp.Chem.* (1998); 19(14); 1639-1662.
47. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; and Taylor, R.; Development and Validation of a Genetic Algorithm for Flexible Docking; *J.Mol.Biol.* (1997); 267; 727-748.
48. Jones, G.; Willett, P.; and Glen, R. C.; Molecular Recognition of Receptor Sites Using a Genetic Algorithm With a Description of Desolvation; *J.Mol.Biol.* (1995 ); 245; 43-53.
49. Judson, R. S.; Tan, Y. T.; Mori, E.; Melius, C.; Jaeger, E. P.; Treasurywalla, A. M.; and Mathiowetz, A.; Docking Flexible Molecules: A Case Study of Three Proteins; *J.Comp.Chem.* (1995); 16 (11); 1405-1419.
50. Payne, A. W. R. and Glen, R. C.; Molecular Recognition Using a Binary Genetic Search Algorithm; *J.Mol.Graphics* (1993); 11; 75-91.
51. von Altrock, C.; Über den Daumen gepeilt; *c't* (1991); 3; 188-206.
52. Rouvray, D. H.; That Fuzzy Feeling in Chemistry; *Chem.Br.* (1995); 31(7); 544-546.
53. "Fuzzy Logic in Chemistry"; Rouvray,D.H.;Ed.; Academic Press; San Diego; 1997.
54. Arkin, A. and Ross, J.; Computational Functions in Biochemical Reaction Networks; *Biophys.J.* (1994); 67; 560-578.
55. Adler, B.; Schütze, P.; and Will, J.; Expert System for Interpretation of X-Ray Diffraction Spectra; *Anal.Chim.Acta.* (1993); 271 ; 287-291.
56. Blaffert, T.; Expertise - an Expert System for Infrared Spectra Evaluation; *Anal.Chim.Acta.* (1986); 191; 161-168.
57. Otto, M.; Fuzzy Logic for Spectra Interpretation; *Anal.Chim.Acta.* (1993); 283; 500-507.
58. Otto, M.; Neural Networks and Fuzzy Logic for Analytical Spectroscopy; *Spec.Publ.-R.Soc.Chem.* (1994); 154; 195-213.
59. Keinan, S. and Avnir, D.; Quantitative Chirality in Structure-Activity Correlations. Shape Recognition by Trypsin, by the D<sub>2</sub> Dopamine Receptor and by Cholinesterases; *J.Am.Chem.Soc.* (1998).
60. Mezey, P. G. and Maruani, J.; The Fundamental Syntopy of Quasi-Symmetric System: Geometric Criteria and the Underlying Syntopy of a Nuclear Configuration Space; *Int.J.Quantum Chem.* (1993); 45(2); 177-187.
61. Mezey, P. G.; Quantum Chemical Shape: New Density Domain Relations for the Topology of Molecular Bodies, Functional Groups, and Chemical Bonding; *Can.J.Chem.* (1994); 72(3); 928-935.
62. Agishtein, M. E.; Fuzzy Molecular Surfaces; *J.Biomol.Struct.& Dynamics* (1992); 9(4); 759-768.

63. Heiden, W. and Brickmann, J.; Segmentation of Protein Surfaces Using Fuzzy Logic; *J.Mol.Graphics* (1994); 12; 106-115.
64. Exner, T. and Brickmann, J.; New Docking Algorithm Based on Fuzzy Set Theory; *J.Mol.Model.* (1997); 3(8); 321-324.
65. Heiden, W.; Methoden zur computerunterstützten Untersuchung selektiver Oberflächeneigenschaften von Proteinen; Dissertation; *Technische Hochschule Darmstadt* (1993).
66. Goldberg, D. E.; "Genetic Algorithms in Search Optimization, and Machine Learning"; Addison-Wesley; MA; 1989.
67. Schöneburg, E.; "Genetische Algorithmen und Evolutionsstrategien. Eine Einführung in Theorie und Praxis der simulierten Evolution."; Addison-Wesley; Bonn; 1994.
68. Nelder, J. A. and Mead, R.; A Simplex Method for Function Minimization; *Computer J.* (1965); 7; 308-313.
69. Levine, D.; "Users Guide to the PGAPack Parallel Genetic Alogrithm Library"; Argonne National Laboratory; Argonne; 1996.
70. Mezey, P. G.; "Molecular Surfaces" in Reviews in Computational Chemistry; 1991; 265-294.
71. Corey, R. and Pauling, L.; Molecular Models of Amino Acids, Peptides and Proteins; *Rev.Sci.Instr.* (1953); 24; 612.
72. Lee, B. and Richards, F. M.; The Interpretation of Protein Structures: Estimation of Static Accessibility; *J.Mol.Biol.* (1971); 55; 379-400.
73. Richards, F. M.; Areas, Volumes, Packing and Protein Structure; *Annu.Rev. Biophys.Bioeng.* (1977); 6; 151-176.
74. Connolly, M. L.; Solvent Accessible Surfaces of Proteins and Nucleic Acids; *Science* (1983); 221; 709-713.
75. Connolly, M. L.; Analytical Molecular Surface Calculation; *J.Appl.Cryst.* (1983); 16; 548-558.
76. Goodford, P.; "Physical Chemistry, Protein Structure and Molecular Design" in Computer-Aided Molecular Design; Richards, G.; Ed.; IBC; London; 1989; 147-155.
77. Hermann, R.; Theory of Hydrophobic Binding. II. The Correlation of Hydrocarbon Solubility in Water With Solvent Cavity Surface Area; *J.Phys.Chem.* (1972); 76; 2754-2759.
78. Bondi, A.; Van der Waals Volumes and Radii; *J.Phys.Chem.* (1964); 68; 441-451.
79. Keil, M.; private Mitteilung (1999).
80. Heiden, W.; Schlenkrich, M.; and Brickmann, J.; Triangulation Algorithms for the Representation of Molecular Surface Properties; *J.Comput.-Aided Mol.Des.* (1990); 4; 255-269.

81. Sanner, M. F.; Olson, A. J.; and Spehner, J.-C.; Reduced Surface: An Efficient Way to Compute Molecular Surfaces; *Biopolymers* (1996); 38; 305-320.
82. Duncan, B. S. and Olson, A. J.; Shape Analysis of Molecular Surfaces; *Biopolymers* (1993); 33; 231-238.
83. Zachmann, C.-D. and Brickmann, J.; Hausdorff Dimension As a Quantification of Local Roughness of Protein Surfaces; *J.Chem.Inf.Comput.Sci.* (1992); 32; 120-122.
84. Spanier, E. H.; "Algebraic Topology"; McGraw-Hill; New York; 1966.
85. Duncan, B. S. and Olson, A. J.; Approximation and Characterization of Molecular Surfaces; *Biopolymers* (1993); 33; 219-229.
86. Zachmann, C.-D.; Topologische Analyse der van der Waals-Oberfläche von Molekülen; Diplomarbeit; *Technische Hochschule Darmstadt* (1990).
87. Zachmann, C.-D.; Heiden, W.; Schlenkrich, M.; and Brickmann, J.; Topological Analysis of Complex Molecular Surfaces; *J.Comp.Chem.* (1992); 1; 76-84.
88. Bronstein, I. N.; Semendjajew, K. A.; Musiol, G.; and Mühlig, H.; "Taschenbuch der Mathematik"; Verlag Harri Deutsch; Frankfurt am Main; 1997.
89. Warwicker, J. and Watson, H. C.; Calculation of the Electric Potential in the Active Site Cleft Due to  $\alpha$ -Helix Dipoles; *J.Mol.Biol.* (1982); 157; 671-679.
90. Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K. A.; and Honig, B. H.; Focusing of Electric Fields in the Active Site of Cu-Zn Superoxide Dismutase: Effects of Ionic Strength and Amino-Acid Modification; *Proteins* (1986); 1; 47-59.
91. Bruccoleri, R. E.; Novotny, J.; Davis, M. E.; and Sharp, K. A.; Finite Difference Poisson-Boltzmann Electrostatic Calculations: Increased Accuracy Achieved by Harmonic Dielectric Smoothing and Charge Antialiasing; *J.Comp.Chem.* (1997); 18(2); 268-276.
92. Nicholls, A. and Honig, B. H.; A Rapid Finite Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson-Boltzmann Equation; *J.Comp.Chem.* (1991); 12(4); 435-445.
93. Gilson, M. K.; Sharp, K. A.; and Honig, B. H.; Calculating the Electrostatic Potential of Molecules in Solution: Method and Error Assessment; *J.Comp.Chem.* (1987); 9(4); 327-335.
94. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; and Karplus, M.; CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations; *J.Comp.Chem.* (1983); 4; 187.
95. Richards, N. G. J.; Williams, P. B.; and Tute, M. S.; Empirical Methods for Computing Molecular Partition Coefficients: II. Inclusion of Conformational Flexibility Within Fragment-Based Approaches; *Int.J.Quant.Chem.* (1992); 44; 219-233.

96. Pixner, P.; Heiden, W.; Merx, H.; Moeckel, G.; Moeller, A.; and Brickmann, J.; Empirical Method for the Quantification and Localization of Molecular Hydrophobicity; *J.Chem.Inf.Comput.Sci.* (1994); 34(6); 1309-1319.
97. Ghose, A. K. and Crippen, G. M.; Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationship I. Partition Coefficients As a Measure of Hydrophobicity; *J.Comp.Chem.* (1986); 7; 565-577.
98. Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; and Robins, R. K.; Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationship. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occuring Nucleoside Antibiotics; *J.Chem.Inf.Comput.Sci.* (1989); 29; 163-172.
99. Ghose, A. K.; Viswanadhan, V. N.; and Wendoloski, J. J.; Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods; *J.Phys.Chem.A* (1998); 102; 3762-3772.
100. Audry, E.; Debost, J. P.; Colleter, J. C.; and Dallet, P.; Une nouvelle approche des relations structure-activité: Le "potentiel de lipophilie moleculaire"; *Eur.J.Med.Chem.* (1986); 21; 71-72.
101. Fauchère, J. L. and Pliska, V.; Hydrophobic Parameters  $\pi$  of Amino-Acid Side Chains From the Partitioning of N-Acetyl-Amino-Acid Amides; *Eur.J.Med.Chem.* (1983); 18; 369-375.
102. Leo, A. J.; Calculating Log P<sub>oct</sub> From Structures; *Chem.Rev.* (1993); 93; 1281-1306.
103. Heiden, W.; Moeckel, G.; and Brickmann, J.; A New Approach to Analysis and Display of Local Lipophilicity/Hydrophilicity Mapped on Molecular Surfaces; *J.Comput.-Aided Mol.Des.* (1993); 7(5); 503-514.
104. Keil, M.; Exner, T. E.; and Brickmann, J.; Characterisation of Protein-Ligand Interfaces: Separating Surface; *J.Mol.Model.* (1998); 4; 335-339.
105. Bernstein, F.; Koetzle, T.; Williams, G.; Meyer, E. jr.; Brice, M.; Rodgers, J.; Kennard, J.; Shimanouchi, T.; and Tasumi, M.; The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures; *J.Mol.Biol.* (1977); 112; 535-542.
106. Fischer, D.; Bachar, O.; Nussinov, R.; and Wolfson, H. J.; An Efficient Automated Computer Vision Based Technique for Detection of Three Dimensional Structural Motifs in Proteins; *J.Biomol.Struct.& Dyn.* (1992); 9(4); 769-789.
107. Fischer, D.; Wolfson, H. J.; and Nussinov, R.; Spatial, Sequence-Order-Independent Structural Comparison of  $\alpha/\beta$  Proteins: Evolutionary Implications; *J.Biomol.Struct.& Dyn.* (1993); 11(2); 367-380.

108. Fischer, D.; Wolfson, H. J.; Lin, S. L.; and Nussinov, R.; Three-Dimensional, Sequence Order-Independent Structural Comparison of a Serine Protease Against the Crystallographic Database Reveals Active Site Similarities: Potential Implications to Evolution and to Protein Folding; *Protein Sci.* (1994); 3; 769-778.
109. Fischer, D.; Tsai, C.-J.; Nussinov, R.; and Wolfson, H. J.; A 3D Sequence-Independent Representation of the Protein Data Bank; *Protein Eng.* (1995); 8(10); 981-997.
110. Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; and Nussinov, R.; A Dataset of Protein-Protein Interfaces Generated With a Sequence-Order-Independent Comparison Technique; *J.Mol.Biol.* (1996); 260; 604-620.
111. Yeates, T. O.; Algorithms for Evaluating the Long-Range Accessibility of Protein Surfaces; *J.Mol.Biol.* (1995); 249; 804-815.
112. Lin, S. L. and Nussinov, R.; Molecular Recognition Via Face Center Representation of Molecular Surface; *J.Mol.Graphics* (1996); 14; 78-90.
113. Ferro, D. R. and Hermans, J.; A Different Best Rigid-Body Molecular Fit Routine; *Acta Cryst.* (1977); A33; 345-347.
114. Späth, H.; "Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion"; R.Oldenbourg Verlag; München, Wien; 1975.
115. Bock, H. H.; "Automatische Klassifikation: Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten"; Vandenhoeck & Ruprecht; Göttingen; 1974.
116. Rarey, M.; Wefing, S.; and Lengauer, Th.; Placement of Medium-Sized Molecular Fragments into Active Sites of Proteins; *J.Comput.-Aided Mol.Des.* (1996); 10; 41-54.
117. Rarey, M.; Kramer, B.; Lengauer, Th.; and Klebe, G.; A Fast Flexible Docking Method Using an Incremental Construction Algorithm; *J.Mol.Biol.* (1996); 261; 470-489.
118. Rarey, M.; Rechnergestützte Vorhersage von Rezeptor-Ligand-Wechselwirkungen; Dissertation; *GMD Sankt Augustin* (1996).
119. Tame, J. R. H.; Scoring Functions: A View From the Bench; *J.Comput.-Aided.Mol.Des.* (1999); 13; 99-108.
120. Böhm, H.-J.; The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein-Ligand Complex of Known Three-Dimensional Structure; *J.Comput.-Aided Mol.Des.* (1994); 8; 243-256.
121. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, Ch. J.; Fogel, D. B.; Fogel, L. J.; and Freer, S. T.; Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming; *Chemistry & Biology* (1995); 2; 317-324.

122. Tomioka, N.; Itai, A.; and Iitaka, Y.; A Method for Fast Energy Estimation and Visualization of Protein-Ligand Interaction; *J.Comput.-Aided Mol.Des.* (1987); 1; 197-210.
123. Meng, E. C.; Kuntz, I. D.; Abraham, D. J.; and Kellogg, G. E.; Evaluating Docked Complexes With the HINT Exponential Function and Empirical Atomic Hydrophobicities; *J.Comput.-Aided Mol.Des.* (1994); 8; 299-306.
124. Jain, A. N.; Scoring Noncovalent Protein-Ligand Interactions: A Continuous Differentiable Function Tuned to Compute Binding Affinities; *J.Comput.-Aided Mol.Des.* (1996); 10; 427-440.
125. Eldridge, M. D.; Murray, Ch. W.; Auton, T. R.; Paolini, G. V.; and Mee, R. P.; Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes; *J.Comput.-Aided Mol.Des.* (1997); 11; 425-445.
126. King, B. L.; Vajda, S.; and DeLisi, C.; Empirical Free Energy As a Target Function in Docking and Design: Application to HIV-1 Protease Inhibitors; *FEBS Letters* (1996); 384; 87-91.
127. Jackson, R. M. and Sternberg, M. J. E.; A Continuum Model for Protein-Protein Interactions: Application to the Docking Problem; *J.Mol.Biol.* (1995); 250; 258-275.
128. Robert, C. H. and Janin, J.; A Soft, Mean-Field Potential Derived From Crystal Contacts for Predicting Protein-Protein Interactions; *J.Mol.Biol.* (1998); 283(5); 1037-1047.
129. Lenhof, H.-P.; New Contact Measures for the Protein Docking Problem; *First Annual International Conference on Computational Molecular Biology* (1997); 1-10.
130. Wallqvist, A.; Jernigan, R. L.; and Covell, D. G.; A Preference-Based Free-Energy Parameterization of Enzyme-Inhibitor Binding. Applications to HIV-1-Protease Inhibitor Design; *Protein Sci.* (1995); 4; 1881-1903.
131. Horton, N. and Lewis, M.; Calculation of the Free Energy of Association for Protein Complexes; *Protein Sci.* (1992); 1; 169-181.
132. Verkhivker, G. M.; Rejto, P. A.; Gehlhaar, D. K.; and Freer, S. T.; Exploring the Energy Landscapes of Molecular Recognition by a Genetic Algorithm: Analysis of the Requirements for Robust Docking of HIV-1 Protease and FKBP-12 Complexes; *Proteins* (1996); 250; 342-353.
133. Press, W. H.; Flannery, B. P.; Tenkolsky, S. A.; and Vetterling, W. T.; "Minimization or Maximization of Functions" in Numerical Recipes in C: The Art of Scientific Computing; Cambridge University Press; Cambridge; 1992; 394-455.

- 134.\* Navia, M. A.; McKeever, B. M.; Springer, J. P.; Lin, T.-Y.; Williams, H. R.; Fluder, E. M.; Dorn, C. P.; and Hoogsteen, K.; Structure of Human Neutrophil Elastase in Complex With a Peptide Chloromethyl Ketone Inhibitor at 1.84-Angstroms Resolution; *Proc.Nat.Acad.Sci.USA* (1989); 86; 7.
- 135.\* Tronrud, D. E.; Monzingo, A. F.; and Matthews, B. W.; Crystallographic Structure Analysis of Phosphoramidates As Inhibitors and Transition-State Analogons of Thermolysin; *Eur.J.Biochem.* (1986); 157; 261.
- 136.\* Cooper, J. B.; Foundling, S. I.; Boger, J.; Blundell, T. L.; Jupp, R. A.; and Kay, J.; X-Ray Studies of Aspartic Proteinase-Statine Inhibitor Complexes; *Biochemistry* (1989); 28; 8596.
- 137.\* Hughes, D. L.; Sieker, L. C.; Bieth, J.; and Dimicoli, J.-L.; Crystallographic Study of the Binding of a Trifluoroacetyl Dipeptide Anilide Inhibitor With Elastase; *J.Mol.Biol.* (1982); 162; 645.
- 138.\* Suguna, K.; Padlan, E. A.; Smith, C. W.; Carlson, W. D.; and Davies, D. R.; Binding of a Reduced Peptide Inhibitor to the Aspartic Proteinase From *Rhizopus Chinensis*. Implications for a Mechanism of Action; *Proc.Nat.Acad.Sci.USA* (1987 ); 84; 7009.
- 139.\* Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; and Kraut, J.; Crystal Structures of *Escherichia Coli* and *Lactobacillus Casei* Dihydrofolate Reductase Refined at 1.7 Angstroms Resolution. I. General Features and Binding of Methotrexate; *J.Biol.Chem.* (1982); 257; 13650.
- 140.\* Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marschall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; and Wlodawer, A.; Structure of Complex of Synthetic HIV-1 Protease With a Substrate-Based Inhibitor at 2.3 Angstroms Resolution; *Science* (1989); 246; 1149.
- 141.\* Suguna, K.; Padlan, E. A.; Bott, R.; Roger, J.; and Davies, D. R.; Structures of Complexes of *Rhizopuspepsin* With Pepstatine and Other Statine-Containing Inhibitors; *to be published* (1989).
- 142.\* James, M. N. G.; Sielecki, A. R.; Brayer, G. D.; Delbaere, L. T. J.; and Bauer, C.-A.; Structures of Product and Inhibitor Complexes of *Streptomyces Griseus* Protease A at 1.8 Angstroms Resolution. A Model for Serine Protease Catalysis; *J.Mol.Biol.* (1980); 43.
- 143.\* Tronrud, D. E.; Holden, H. M.; and Matthews, B. W.; Structures of Two Thermolysin-Inhibitor Complexes That Differ by a Single Hydrogen Bond; *Science* (1987); 235; 571.
- 144.\* Frigerio, F.; Coda, A.; Pugliese, L.; Lionetti, C.; Menegatti, E.; Amiconi, G.; Schnebli, H. P.; Ascenzi, P.; and Bolognesi, M.; Crystal and Molecular Structure of the Bovine Alpha-Chymotrypsin-Eglin C Complex at 2.0 Angstroms Resolution; *J.Mol.Biol.* (1992); 225; 107.



- 145.\* Scheidig, A. J.; Hynes, T. R.; Pelletier, L. A.; and Kossiakoff, A. A.; Crystal Structures of Bovine Chymotrypsin and Trypsin Complexed to the Inhibitor Domain of Alzheimer's B-Protein Precursor (Appi) and Basic Pancreatic Trypsin Inhibitor (Bpti): Examples of the Binding Modes of Non-Optimum P1 Residues; *to be published* (1996).
- 146.\* Fujinaga, M.; Sielecki, A. R.; Read, R. J.; Ardelt, W.; Laskowski Jr., M.; and James, M. N. G.; Crystal and Molecular Structures of the Complex of Alpha-Chymotrypsin With Its Inhibitor Turkey Ovomucoid Third Domain at 1.8 Angstroms Resolution; *J.Mol.Biol.* (1987); 195; 397.
- 147.\* Bode, W.; Papamokos, E.; and Musil, D.; The High-Resolution X-Ray Crystal Structure of the Complex Formed Between Subtilisin Carlsberg and Eglin C, an Elastase Inhibitor From the Leech *Hirudo Medicinalis*. Structure Analysis, Subtilisin Structure and Interface Geometry; *Eur.J.Biochem.* (1987); 166; 674.
- 148.\* Bode, W.; Wei, A. Z.; Huber, R.; Meyer, E.; Travis, J.; and Neumann, S.; X-Ray Crystal Structure of the Complex of Human Leukocyte Elastase (Pmn Elastase) and the Third Domain of the Turkey Ovomucoid Inhibitor; *Embo J.* (1986); 5; 2453.
- 149.\* Gros, P.; Fujinaga, M.; Dijkstra, B. W.; Kalk, K. H.; and Hol, W. G. J.; Crystallographic Refinement by Incorporation of Molecular Dynamics. The Thermostable Serine Protease Thermitase Complexed With Eglin-C; *to be published* (1989).
- 150.\* Bolognesi, M.; Gatti, G.; Menegatti, E.; Guarneri, M.; Marquart, M.; Papamokos, E.; and Huber, R.; Three-Dimensional Structure of the Complex Between Pancreatic Secretory Inhibitor (Kazal Type) and Trypsinogen at 1.8 Angstroms Resolution. Structure Solution, Crystallographic Refinement and Preliminary Structural Interpretation; *J.Mol.Biol.* (1982); 162; 839.
- 151.\* Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; and Huber, R.; The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and Its Complexes With Inhibitors; *Acta Crystallogr.,Sect.B* (1983); 39; 480.
- 152.\* Knighton, D. R.; Zheng, J.; Ten Eyck, L. F.; Xuong, V. A.; Ashford, N.-H.; Taylor, S. S.; and Sowadski, J. M.; Crystal Structure of the Catalytic Subunit of Cyclic Adenosine Monophosphate-Dependent Protein Kinase; *Science* (1991); 253; 407.
- 153.\* Chen, Z. and Bode, W.; Refined 2.5 Angstroms X-Ray Crystal Structure of the Complex Formed by Porcine Kallikrein A and the Bovine Pancreatic Trypsin Inhibitor. Crystallization, Patterson Search, Structure Determination, Refinement, Structure and Comparison With Its Components and With the Bovine Trypsin-Pancreatic Trypsin Inhibitor Complex; *J.Mol.Biol.* (1983); 164; 283.

- 154.\* McPhalen, C. A. and James, M. N. G.; Structural Comparison of Two Serine Proteinase-Protein Inhibitor Complexes. Eglin-C-Subtilisin Carlsberg and /Ci-2-Subtilisin Novo; *Biochemistry* (1988); 27; 6582.
- 155.\* Takeuchi, Y.; Satow, Y.; Nakamura, K. T.; and Mitsui, Y.; Refined Crystal Structure of the Complex of Subtilisin/BPN' and Streptomyces Subtilisin Inhibitor at 1.8 Angstroms Resolution; *J.Mol.Biol.* (1991); 221; 309.
- 156.\* Rees, D. C. and Lipscomb, W. N.; Refined Crystal Structure of the Potato Inhibitor Complex of Carboxypeptidase A at 2.5 Angstroms Resolution; *J.Mol.Biol.* (1982 ); 160; 475.
- 157.\* Greenblatt, H. M.; Rayn, C. A.; and James, M. N. G.; Structure of the Complex of Streptomyces Griseus Proteinase B and Polypeptide Chymotrypsin Inhibitor-1 From Russet Burbank Potato Tubers at 2.1 Angstroms Resolution; *J.Mol.Biol.* (1989); 205; 201.
- 158.\* Bode, W.; Walter, J.; Huber, R.; Wenzel, H. R.; and Tschesche, H.; The Refined 2.2-Angstroms (0.22-Nm) X-Ray Crystal Structure of the Ternary Complex Formed by Bovine Trypsinogen, Valine-Valine and the Arg=15=Analogue of Bovine Pancreatic Trypsin Inhibitor; *Eur.J.Biochem.* (1984); 144; 185.
- 159.\* Cook, W. J.; Jeffrey, L. C.; Carson, M.; Chen, Z.; and Pickart, C. M.; Structure of a Diubiquitin Conjugate and a Model for Interaction With Ubiquitin Conjugating Enzyme (E2); *J.Biol.Chem.* (1992); 267; 16467.
- 160.\* Finzel, B. C.; Weber, P. C.; Hardman, K. D.; and Salemme, F. R.; Structure of Ferricytochrome C (Prime) From Rhodospirillum Molischianum at 1.67 Angstroms Resolution; *J.Mol.Biol.* (1985); 186; 627.
- 161.\* Bally, R. and Delettre, J.; Structure and Refinement of the Oxidized P2=1=Form of Uteroglobin at 1.64 Angstroms Resolution; *J.Mol.Biol.* (1989); 206; 153.
- 162.\* Fischmann, T. O.; Bentley, G. A.; Bhat, T. N.; Boulot, G.; Mariuzza, R. A.; Phillips, S. E. V.; Tello, D.; and Poljak, R. J.; Crystallographic Refinement of the Three-Dimensional Structure of the Fab 1.3-Lysozyme Complex at 2.5-Anstroms Resolution; *J.Biol.Chem.* (1991); 266; 12915.
- 163.\* Cohen, G. H.; Sheriff, S.; and Davies, D. R.; The Refined Structure of the Monoclonal Antibody Hy/Hel-5 With Its Antigen Hen Egg White Lysozyme; *to be published* (1995).
- 164.\* Padlan, E. A.; Silverton, E. W.; Scheriff, S.; Cohen, G. H.; Smith-Gill, S. J.; and Davies, D. R.; Structure of an Antibody-Antigen Complex. Crystal Structure of the Hy/Hel-10 Fab-Lysozyme Complex; *Proc.Nat.Acad.Sci.USA* (1989); 86; 5938.
- 165.\* Stanfield, R. L.; Fieser, T. M.; Lerner, R. A.; and Wilson, I. A.; Crystal Structures of an Antibody to a Peptide and Its Complex With Peptide Antigen at 2.8 Angstroms; *Science* (1990); 248; 712.

166. Keil, M.; Computergestützte Untersuchung der Bindungsregion des p53-DNA-Komplexes; Diplomarbeit; *Technische Hochschule Darmstadt* (1996).
167. SYBYL 6.6, Tripos Inc., 1699 South Hanley Road, St. Louis, Missouri, 63144, USA.
168. Jäger, R.; Empirische Berechnungsverfahren von oberflächenbezogenen 1-Oktanol / Wasser-Verteilungskoeffizienten; Diplomarbeit; *Technische Hochschule Darmstadt* (1996).
169. Pirard, B.; Baudoux, G.; and Durant, F.; A Database Study of Intermolecular NH...O Hydrogen Bonds for Carboxylates, Sulfonates and Monohydrogen Phosphonates; *Acta Cryst.* ( 1995); *B51*; 103-107.
170. Shoichet, B. K. and Kuntz, I. D.; Predicting the Structure of Protein Complexes: a Step in the Right Direction; *Chemistry & Biology* (1996); 3; 151-156.
171. Dixon, J. S.; Evaluation of the CASP2 Docking Section; *Proteins* (1997); *Suppl 1*; 198-204.
172. Judson, R.; "Genetic Algorithms and Their Use in Chemistry" in Reviews in Computational Chemistry; Lipkowitz, K. B. and Boyd, D. B.; Eds.; VCH Publishers, Inc.; New York; 1997; 1-73.
173. Jäger, R.; Dissertation; *Technische Universität Darmstadt*, in Vorbereitung.
174. Judson, R. S. and Rabitz, H.; Teaching Lasers to Control Molecules; *Phys.Rev.Lett.* (1992); 68(10); 1500-1503.
175. Clark, M.; Cramer, R. D., III; and Van Opdenbosch, N.; Validation of the General Purpose Tripos 5.2 Force Field; *J.Comp.Chem.* (1989); 10(8); 982-1012.

\* Die Literaturstellen wurden ohne Änderungen der *Brookhaven Protein Data Bank* (<http://www.rcsb.org/pdb/index.html>) entnommen.

## Anhang 1: Domänen der Enzym-Inhibitor-Komplexe mit niedermolekularen Inhibitoren

Tabelle A1.1: Anzahl der nach topographischen Kriterien eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Topographie

<i>pdb</i> -Datei	gesamt	Pfropf	Grat	Sattel	Spalt	Loch
<b>1hne</b>	689	117	168	192	132	80
	30	10	12	6	2	0
<b>1tlp</b>	933	158	215	258	182	120
	30	10	12	6	2	0
<b>2er9</b>	928	164	213	251	185	115
	64	18	22	16	7	1
<b>2est</b>	738	123	177	210	144	84
	23	6	11	5	1	0
<b>3apr</b>	941	151	226	266	189	109
	63	17	22	15	6	3
<b>4dfr</b>	607	98	144	170	118	77
	27	9	12	6	0	0
<b>4hvp</b>	690	123	170	186	132	79
	55	15	19	13	6	2
<b>5apr</b>	947	151	227	257	200	112
	47	14	18	10	4	1
<b>5sga</b>	567	94	141	160	112	60
	30	10	11	6	3	0
<b>6tmn</b>	944	159	217	265	181	122
	30	10	11	7	2	0

Tabelle A1.2: Anzahl der nach dem elektrostatischen Potential eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Elektrostatik  
Die Berechnung des elektrostatischen Potentials erfolgte mit dem Poisson-Boltzmann-Ansatz.

<i>pdb</i> -Datei	gesamt	stark neg.	negativ	neutral	positiv	stark pos.
<b>1hne</b>	406	34	84	175	113	0
	32	6	8	13	5	0
<b>1tlp</b>	338	7	65	249	17	0
	21	3	8	10	0	0
<b>2er9</b>	326	3	58	263	2	0
	46	3	11	24	8	0
<b>2est</b>	270	1	30	209	30	0
	23	4	3	9	7	0
<b>3apr</b>	364	8	82	258	16	0
	54	4	14	26	10	0
<b>4dfr</b>	296	9	86	158	43	0
	18	2	6	9	1	0
<b>4hvp</b>	277	1	42	195	39	0
	46	6	9	21	10	0
<b>5apr</b>	310	3	34	263	10	0
	35	2	8	19	6	0
<b>5sga</b>	267	3	58	173	33	0
	28	4	8	12	4	0
<b>6tmn</b>	362	6	63	253	40	0
	20	3	6	11	0	0

Tabelle A1.3: Anzahl der nach dem elektrostatischen Potential eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Elektrostatik  
Die Berechnung des elektrostatischen Potentials erfolgte mit dem modifizierten Coulomb-Ansatz.

<i>pdb</i> -Datei	gesamt	stark neg.	negativ	neutral	positiv	stark pos.
<b>1hne</b>	322	39	51	127	105	0
	32	7	7	11	7	0
<b>1tlp</b>	381	15	100	207	59	0
	20	5	9	6	0	0
<b>2er9</b>	392	30	103	209	50	0
	52	10	12	17	13	0
<b>2est</b>	355	27	72	162	94	0
	15	0	1	7	5	2
<b>3apr</b>	412	20	115	211	66	0
	38	4	10	17	7	0
<b>4dfr</b>	296	49	81	110	56	0
	17	5	5	7	0	0
<b>4hvp</b>	318	23	64	156	75	0
	34	0	6	14	10	4
<b>5apr</b>	467	35	137	210	85	0
	38	5	8	17	8	0
<b>5sga</b>	253	7	48	146	52	0
	22	2	8	11	1	0
<b>6tmn</b>	385	15	86	218	66	0
	18	4	6	8	0	0

Tabelle A1.4: Anzahl der nach der Lipophilie eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Lipophilie

<i>pdb</i> -Datei	gesamt	hydrophil	neutral	lipophil
<b>1hne</b>	279	127	129	23
	19	11	8	0
<b>1tlp</b>	313	129	170	14
	19	7	8	4
<b>2er9</b>	309	123	184	2
	30	9	10	11
<b>2est</b>	242	77	160	5
	14	2	8	4
<b>3apr</b>	337	159	168	10
	36	9	18	9
<b>4dfr</b>	195	69	125	1
	14	4	9	1
<b>4hvp</b>	254	96	140	14
	29	11	13	5
<b>5apr</b>	332	149	172	11
	31	10	14	7
<b>5sga</b>	217	118	89	10
	17	5	10	2
<b>6tmn</b>	313	126	175	12
	18	3	9	6

## Anhang 2: Domänen der Enzym-Inhibitor-Komplexe mit inhibierenden Proteinen

Tabelle A2.1: Anzahl der nach topographischen Kriterien eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Topographie

<i>pdb</i> -Datei	gesamt	Pfropf	Grat	Sattel	Spalt	Loch
<b>1acb</b>	444	82	103	124	84	51
	170	35	42	48	31	14
<b>1cbw</b>	447	80	104	120	87	56
	175	38	45	50	28	14
<b>1cho</b>	442	74	95	119	86	48
	151	31	41	43	25	11
<b>1cpk</b>	680	119	156	188	127	90
	81	22	24	19	12	4
<b>1cse</b>	465	85	98	129	93	60
	176	38	43	47	32	16
<b>1ppf</b>	425	78	99	126	78	44
	160	32	41	48	26	13
<b>1tec</b>	440	83	97	129	81	50
	176	37	43	50	34	12
<b>1tgs</b>	411	81	91	120	74	45
	176	35	48	49	29	15
<b>1tpa</b>	416	78	95	117	82	44
	167	36	42	45	29	15
<b>2cpk</b>	680	119	156	188	127	90
	81	22	24	19	12	4
<b>2kai</b>	441	77	104	126	84	50
	149	31	38	41	28	11
<b>2ptc</b>	407	76	90	121	79	41
	161	32	41	48	26	14
<b>2sec</b>	443	87	101	120	82	53
	186	39	44	58	30	15
<b>2sic</b>	460	89	105	127	91	48
	245	51	57	68	44	25
<b>2sni</b>	449	87	104	120	82	56
	176	35	40	53	29	19
<b>2tgp</b>	416	78	92	115	81	50
	172	34	45	45	31	17
<b>4cpa</b>	540	99	121	144	117	59
	120	25	32	35	17	11
<b>4sgb</b>	375	73	89	104	70	39
	132	29	36	38	20	9
<b>4tpi</b>	400	76	90	115	78	41
	178	36	45	49	31	17



Tabelle A2.2: Anzahl der nach dem elektrostatischen Potential eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Elektrostatik  
Die Berechnung des elektrostatischen Potentials erfolgte mit dem modifizierten Coulomb-Ansatz.

<i>pdb</i> -Datei	gesamt	stark neg.	negativ	neutral	positiv	stark pos.
<b>1acb</b>	238	33	57	90	58	0
	90	11	24	35	20	0
<b>1cbw</b>	226	31	59	87	49	0
	103	19	17	34	33	0
<b>1cho</b>	217	31	51	81	54	0
	92	15	24	34	19	0
<b>1cpk</b>	364	53	83	138	90	0
	50	10	10	17	13	0
<b>1cse</b>	196	12	43	95	46	0
	89	10	21	37	21	0
<b>1ppf</b>	202	19	37	85	61	0
	97	16	24	36	21	0
<b>1tec</b>	213	17	46	91	59	0
	91	11	25	37	18	0
<b>1tgs</b>	182	16	32	85	49	0
	105	10	23	43	29	0
<b>1tpa</b>	208	20	46	81	61	0
	103	19	18	37	29	0
<b>2cpk</b>	364	53	83	138	90	0
	50	10	10	17	13	0
<b>2kai</b>	204	22	68	92	22	0
	82	16	17	27	22	0
<b>2ptc</b>	195	20	36	85	54	0
	86	16	13	30	27	0
<b>2sec</b>	162	3	26	99	34	0
	89	11	22	36	20	0
<b>2sic</b>	192	12	41	98	41	0
	122	16	36	48	22	0
<b>2sni</b>	188	12	41	92	43	0
	94	15	25	37	17	0
<b>2tgp</b>	212	20	38	92	62	0
	100	21	17	33	29	0
<b>4cpa</b>	174	2	26	116	30	0
	72	9	20	28	15	0
<b>4sgb</b>	168	9	36	82	41	0
	82	9	20	29	24	0
<b>4tpi</b>	206	23	43	82	58	0
	97	16	15	34	32	0

Tabelle A2.3: Anzahl der nach der Lipophilie eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Lipophilie

<i>pdb</i> -Datei	gesamt	hydrophil	neutral	lipophil
<b>1acb</b>	160	66	91	3
	58	16	39	3
<b>1cbw</b>	159	68	82	9
	68	24	39	5
<b>1cho</b>	144	58	84	2
	60	24	34	2
<b>1cpk</b>	221	81	137	3
	33	14	17	2
<b>1cse</b>	148	62	84	2
	62	19	37	6
<b>1ppf</b>	172	70	87	15
	62	23	38	1
<b>1tec</b>	156	73	71	12
	67	21	38	8
<b>1tgs</b>	144	58	80	6
	63	22	40	1
<b>1tpa</b>	138	59	73	6
	70	26	40	4
<b>2cpk</b>	221	81	137	3
	33	14	17	2
<b>2kai</b>	151	50	97	4
	50	15	34	1
<b>2ptc</b>	142	61	75	6
	57	20	34	3
<b>2sec</b>	147	61	82	4
	63	19	37	7
<b>2sic</b>	158	69	86	3
	84	29	53	2
<b>2sni</b>	158	72	80	6
	69	24	39	6
<b>2tgp</b>	141	60	79	2
	66	21	43	2
<b>4cpa</b>	178	62	107	9
	48	19	26	3
<b>4sgb</b>	124	62	58	4
	50	17	29	4
<b>4tpi</b>	147	61	82	4
	71	24	44	3

### Anhang 3: Domänen der Protein-Dimere

Tabelle A3.1: Anzahl der nach topographischen Kriterien eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Topographie

<i>pdb</i> -Datei	gesamt	Pfropf	Grat	Sattel	Spalt	Loch
<b>1aar</b>	227	47	56	66	36	22
	220	48	52	63	38	19
<b>2ccy</b>	324	62	73	90	62	37
	316	56	71	95	59	35
<b>2utg</b>	220	46	50	59	40	25
	232	46	53	64	42	27

Tabelle A3.2: Anzahl der nach dem elektrostatischen Potential eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Elektrostatik  
Die Berechnung des elektrostatischen Potentials erfolgte mit dem modifizierten Coulomb-Ansatz.

<i>pdb</i> -Datei	gesamt	stark neg.	negativ	neutral	positiv	stark pos.
<b>1aar</b>	117	20	26	45	26	0
	118	20	25	45	28	0
<b>2ccy</b>	143	19	34	62	28	0
	149	22	37	61	29	0
<b>2utg</b>	106	18	25	39	24	0
	95	15	22	37	21	0

Tabelle A3.3: Anzahl der nach der Lipophilie eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Lipophilie

<i>pdb</i> -Datei	gesamt	hydrophil	neutral	lipophil
<b>1aar</b>	72	28	42	2
	70	20	50	0
<b>2ccy</b>	103	39	62	2
	104	41	61	2
<b>2utg</b>	73	24	43	6
	75	25	45	5

## Anhang 4: Domänen der Antikörper-Antigen-Komplexe

Tabelle A4.1: Anzahl der nach topographischen Kriterien eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Topographie

<i>pdb</i> -Datei	gesamt	Pfropf	Grat	Sattel	Spalt	Loch
<b>1fdl</b>	772	139	176	227	146	84
	252	45	65	71	48	23
<b>3hfl</b>	882	155	192	247	181	107
	281	48	67	83	54	29
<b>3hfm</b>	921	161	196	256	181	127
	292	54	71	86	51	30
<b>2igf</b>	1422	233	328	388	282	191
	49	14	19	11	4	1

Tabelle A4.2: Anzahl der nach dem elektrostatischen Potential eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Elektrostatik  
Die Berechnung des elektrostatischen Potentials erfolgte mit dem modifizierten Coulomb-Ansatz.

<i>pdb</i> -Datei	gesamt	stark neg.	negativ	neutral	positiv	stark pos.
<b>1fdl</b>	466	52	117	185	112	0
	141	20	29	52	40	0
<b>3hfl</b>	412	42	105	184	81	0
	148	23	29	51	45	0
<b>3hfm</b>	306	7	59	188	52	0
	149	22	25	53	49	0
<b>2igf</b>	684	55	160	338	131	0
	44	8	7	16	13	0

Tabelle A4.3: Anzahl der nach der Lipophilie eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Lipophilie

<i>pdb</i> -Datei	gesamt	hydrophil	neutral	lipophil
<b>1fdl</b>	292	131	160	1
	108	49	48	11
<b>3hfl</b>	300	130	166	4
	104	48	46	10
<b>3hfm</b>	295	125	167	3
	109	52	48	9
<b>2igf</b>	468	187	276	5
	23	9	14	0

## Anhang 5: *Docking*-Ergebnisse für Enzym-Ligand-Komplexe mit inhibierenden Proteinen, Protein-Dimere und Antikörper-Antigen-Komplexe

Tabelle A5.1: *Docking*-Vorhersagen durch Vergleich der segmentierten Oberflächen (modifizierter Coulomb-Ansatz)

<i>pdb</i> -Datei	Anzahl komplementärer Domänen <sup>1)</sup>	Anzahl Strukturen	Anzahl von komplementären Domänen der besten Struktur	<i>rms</i> -Wert (Å)	Rechenzeit <sup>2)</sup> (min:s)
<b>1acb</b>	39149 (19363, 13436, 6350)	60522	15	3.840	17:56
<b>1cbw</b>	41462 (20393, 13569, 7500)	99298	21	1.083	25:47
<b>1cho</b>	34675 (16118, 11910, 6647)	50355	19	1.994	12:25
<b>1cse</b>	41106 (21677, 13015, 6414)	86819	15	2.145	22:50
<b>1ppf</b>	36689 (15864, 12793, 8032)	50525	14	1.716	12:15
<b>1tec</b>	41090 (21137, 13600, 6353)	84139	30	0.904	20:18
<b>1tgs</b>	38144 (17628, 13532, 6984)	55759	23	3.062	13:37
<b>1tpa</b>	35008 (16565, 11676, 6767)	41544	14	3.543	11:26
<b>2cpk</b>	27833 (12512, 10259, 5062)	20759	11	4.177	8:11
<b>2kai</b>	34023 (16562, 11673, 5788)	42213	25	2.462	12:30
<b>2ptc</b>	31575 (16416, 9826, 5333)	55795	25	3.651	15:17
<b>2sec</b>	39148 (21093, 12073, 5982)	56500	27	0.858	14:30
<b>2sic</b>	57368 (29960, 17150, 10258)	151653	11	2.566	39:28
<b>2sni</b>	40602 (20669, 12581, 7352)	63252	24	1.338	17:00
<b>2tgp</b>	37182 (17631, 12523, 7028)	58262	29	3.206	14:44
<b>4cpa</b>	32810 (16403, 10554, 5853)	55343	16	4.481	13:45
<b>4sgb</b>	25954 (11652, 9966, 4336)	23638	15	3.655	5:40
<b>4tpi</b>	36470 (17363, 11280, 7827)	43634	27	1.514	11:47
<b>1aar</b>	23429 (11632, 7774, 4023)	34870	17	0.971	8:07
<b>2ccy</b>	48573 (27018, 13201, 8354)	138110	38	0.498	41:23
<b>2utg</b>	22220 (12777, 5928, 3515)	51654	38	1.354	12:58

Fortsetzung Tabelle A5.1

<i>pdb</i> -Datei	Anzahl komplementärer Domänen <sup>1)</sup>	Anzahl Strukturen	Anzahl von komplementären Domänen der besten Struktur	<i>rms</i> -Wert (Å)	Rechenzeit <sup>2)</sup> (min:s)
<b>1fdl</b>	121889 (56763, 42286, 22840)	487603	19	4.280	151:27
<b>3hfl</b>	132409 (70005, 40442, 21962)	574705	27	4.785	168:10
<b>3hfm</b>	134950 (75658, 35925, 23367)	796912	34	2.012	215:33
<b>2igf</b>	39349 (11774, 19441, 8134)	22402	27	1.373	7:02

1) Anzahl der Domänenpaare mit einer Komplementarität über 0.8

Die erste Zahl ist die Gesamtzahl aller Domänenpaare, in Klammern sind die nach topographischen Kriterien, nach dem elektrostatischen Potential und nach der Lipophilie eingeteilten Domänen getrennt aufgeführt. Für die Berechnung des elektrostatischen Potentials wurde der modifizierte Coulomb-Ansatz verwendet.

2) Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor

Tabelle A5.2: Ergebnisse des Tests auf Volumenüberlappung und des Clusteralgorithmus (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	Anzahl der Strukturen	<i>rms</i> -Wert (Å) des besten Strukturs	Rechenzeit (h:min)	Anzahl der Cluster	<i>rms</i> -Wert (Å) des besten Clusters	Rechenzeit (s)
<b>1acb</b>	16489	3.840	0:12	4604	4.276	23.09
<b>1cbw</b>	21810	1.083	0:18	5222	2.515	30.49
<b>1cho</b>	10046	1.994	0:09	2865	3.115	8.21
<b>1cse</b>	30774	2.145	0:15	6801	3.250	58.06
<b>1ppf</b>	9628	1.716	0:10	2720	3.718	7.13
<b>1tec</b>	28598	0.904	0:16	6305	2.231	51.20
<b>1tgs</b>	13584	3.062	0:10	3725	3.062	13.08
<b>1tpa</b>	11799	3.543	0:07	3271	3.543	12.23
<b>2cpk</b>	3564	4.421	0:03	1244	4.421	1.59
<b>2kai</b>	7650	2.462	0:09	2596	3.056	6.69
<b>2ptc</b>	15585	3.651	0:10	3714	4.137	17.97
<b>2sec</b>	18752	0.858	0:10	4915	1.110	25.10
<b>2sic</b>	44936	2.566	0:37	10389	2.566	87.74
<b>2sni</b>	17098	1.338	0:12	4678	1.130	20.38
<b>2tgp</b>	16685	3.206	0:10	4155	3.439	22.17
<b>4cpa</b>	21183	4.481	0:07	3865	5.642	27.77
<b>4sgb</b>	8160	3.655	0:03	2138	5.142	6.29
<b>4tpi</b>	12326	1.514	0:08	3303	3.550	13.34
<b>1aar</b>	14573	0.971	0:05	3547	2.836	22.41
<b>2ccy</b>	30272	0.498	0:38	9376	1.255	60.74
<b>2utg</b>	15604	1.354	0:10	3525	1.646	21.48
<b>1fdl</b>	40063	4.701	4:52	13500	5.770	96.77
<b>3hfl</b>	59025	4.785	5:22	18156	4.785	149.88
<b>3hfm</b>	86856	2.012	7:46	22859	2.213	250.07
<b>2igf</b>				6499	1.237	35.13

Tabelle A5.3: Ergebnisse der rigiden Energieminimierung mit dem *Downhill Simplex*-Algorithmus (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	<i>rms</i> -Wert (Å) der besten Struktur	Rang der besten Struktur	Energie der besten Struktur	Rechenzeit (h:min)
1acb	0.895	1	-167	1:21
1cbw	0.467	1	-167	1:24
1cho	0.538	1	-177	0:39
1cse	0.228	1	-192	2:03
1ppf	0.474	1	-161	0:37
1tec	1.056	1	-183	1:53
1tgs	0.736	1	-214	0:51
1tpa	0.900	1	-199	0:47
2cpk	0.507	1	-228	0:10
2kai	0.699	1	-178	0:40
2ptc	0.642	1	-195	0:54
2sec	0.417	1	-187	1:32
2sic	0.855	1	-197	4:23
2sni	0.211	1	-177	1:32
2tgp	0.588	1	-187	1:00
4cpa	2.535	2	-99	0:46
4sgb	0.529	1	-165	0:24
4tpi	0.621	1	-207	0:48
1aar	0.985	1	-116	0:35
2ccy	0.430	1	-157	3:07
2utg	0.669	1	-221	0:33
1fdl	0.584	1	-147	11:14
3hfl	0.879	1	-194	14:56
3hfm	0.494	1	-161	18:39
2igf	0.387	1	-147	0:23

Tabelle A5.4: Ergebnisse des *Clustering* der minimierten Komplexstrukturen (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	Anzahl der Cluster	<i>rms</i> -Wert (Å) des besten Clusters	Rechenzeit (s)
<b>1acb</b>	4354	1.190	2.15
<b>1cbw</b>	4926	0.523	2.44
<b>1cho</b>	2778	0.538	1.11
<b>1cse</b>	6332	0.834	3.56
<b>1ppf</b>	2617	1.176	0.99
<b>1tec</b>	5927	0.634	3.65
<b>1tgs</b>	3532	0.678	1.38
<b>1tpa</b>	3112	0.650	1.41
<b>2cpk</b>	1188	0.526	0.36
<b>2kai</b>	2475	0.600	0.97
<b>2ptc</b>	3494	0.679	1.61
<b>2sec</b>	4624	0.736	2.19
<b>2sic</b>	9472	1.388	6.17
<b>2sni</b>	4427	0.236	1.99
<b>2tgp</b>	3942	0.705	1.86
<b>4cpa</b>	3691	2.535	1.80
<b>4sgb</b>	2015	0.453	0.76
<b>4tpi</b>	3133	0.455	1.31
<b>1aar</b>	3372	0.985	1.90
<b>2ccy</b>	8670	0.360	5.45
<b>2utg</b>	3327	0.697	1.59
<b>1fdl</b>	12887	0.704	4.07
<b>3hfl</b>	17258	0.876	5.70
<b>3hfm</b>	22859	0.471	7.98
<b>2igf</b>	4512	0.354	4.35



## Anhang 6: Ergebnisse des flexiblen *Dockings* von Methotrexat in Dihydrofolat-Reduktase

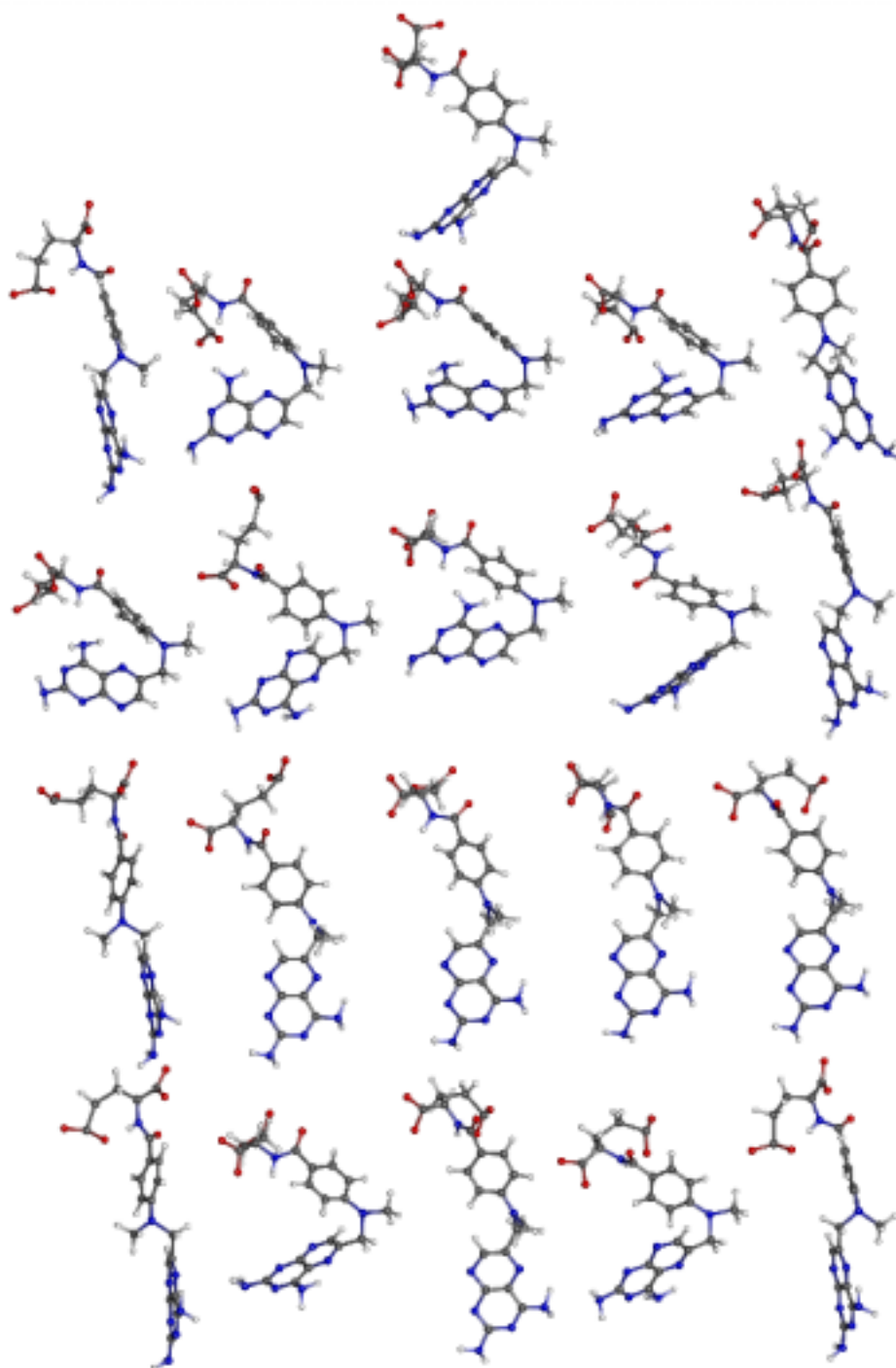


Abb. A6.1: Ausgangskonfigurationen des Methotrexat für das flexible *Docking* mit Dihydrofolat-Reduktase. Zum Vergleich ist zuerst die im Komplex kristallographisch bestimmte Struktur gegeben. Darauf folgen die zwanzig durch den *GA Conformational Search*-Algorithmus bestimmten niederenergetischen Konformationen

Tabelle A6.1: Anzahl der nach topographischen Kriterien eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Topographie

<i>pdb</i> -Datei	gesamt	Pfropf	Grat	Sattel	Spalt	Loch
<b>Enzym</b>	607	98	144	170	118	77
<b>conf1</b>	25	8	12	5	0	0
<b>conf2</b>	26	8	11	5	2	0
<b>conf3</b>	31	9	10	7	3	2
<b>conf4</b>	27	8	11	7	1	0
<b>conf5</b>	25	8	12	5	0	0
<b>conf6</b>	32	9	11	7	3	2
<b>conf7</b>	27	9	12	6	0	0
<b>conf8</b>	30	7	13	8	2	0
<b>conf9</b>	24	9	10	5	0	0
<b>conf10</b>	22	6	12	4	0	0
<b>conf11</b>	23	6	12	5	0	0
<b>conf12</b>	24	7	13	4	0	0
<b>conf13</b>	22	7	12	3	0	0
<b>conf14</b>	27	10	11	6	0	0
<b>conf15</b>	24	7	11	6	0	0
<b>conf16</b>	28	11	12	5	0	0
<b>conf17</b>	26	7	12	6	1	0
<b>conf18</b>	24	7	12	5	0	0
<b>conf19</b>	25	7	11	7	0	0
<b>conf20</b>	25	8	11	6	0	0

Tabelle A6.2: Anzahl der nach dem elektrostatischen Potential eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Elektrostatik  
Die Berechnung des elektrostatischen Potentials erfolgte mit dem modifizierten Coulomb-Ansatz.

<i>pdb</i> -Datei	gesamt	stark neg.	negativ	neutral	positiv	stark pos.
<b>Enzym</b>	296	49	81	110	56	0
<b>conf1</b>	15	5	5	5	0	0
<b>conf2</b>	15	5	7	3	0	0
<b>conf3</b>	17	5	5	6	1	0
<b>conf4</b>	15	5	6	4	0	0
<b>conf5</b>	18	6	5	7	0	0
<b>conf6</b>	17	5	5	6	1	0
<b>conf7</b>	13	5	3	5	0	0
<b>conf8</b>	15	4	5	5	1	0
<b>conf9</b>	15	5	4	6	0	0
<b>conf10</b>	14	4	4	6	0	0
<b>conf11</b>	15	4	4	7	0	0
<b>conf12</b>	16	5	4	7	0	0
<b>conf13</b>	15	5	4	6	0	0
<b>conf14</b>	18	5	5	8	0	0
<b>conf15</b>	16	5	5	6	0	0
<b>conf16</b>	16	5	5	6	0	0
<b>conf17</b>	13	5	3	5	0	0
<b>conf18</b>	17	5	5	7	0	0
<b>conf19</b>	15	5	4	6	0	0
<b>conf20</b>	17	6	5	6	0	0

Tabelle A6.3: Anzahl der nach der Lipophilie eingeteilten Domänen und die Zuordnung zu den einzelnen Klassen der linguistischen Variablen Lipophilie

<i>pdb</i> -Datei	gesamt	hydrophil	neutral	lipophil
<b>Enzym</b>	195	69	125	1
<b>conf1</b>	16	4	11	1
<b>conf2</b>	14	5	8	1
<b>conf3</b>	12	4	7	1
<b>conf4</b>	13	3	9	1
<b>conf5</b>	14	4	9	1
<b>conf6</b>	13	4	8	1
<b>conf7</b>	13	4	8	1
<b>conf8</b>	12	4	7	1
<b>conf9</b>	13	4	8	1
<b>conf10</b>	14	4	9	1
<b>conf11</b>	14	5	8	1
<b>conf12</b>	13	4	8	1
<b>conf13</b>	14	4	9	1
<b>conf14</b>	15	4	9	2
<b>conf15</b>	14	4	9	1
<b>conf16</b>	13	4	8	1
<b>conf17</b>	14	4	9	1
<b>conf18</b>	14	4	8	2
<b>conf19</b>	16	4	11	1
<b>conf20</b>	13	4	8	1

Tabelle A6.4: *Docking*-Vorhersagen durch Vergleich der segmentierten Oberflächen (modifizierter Coulomb-Ansatz)

<i>pdb</i> -Datei	Anzahl komplementärer Domänen <sup>1)</sup>	Anzahl Strukturen	Anzahl von komplementären Domänen der besten Struktur	<i>rms</i> -Wert (Å)	Rechenzeit <sup>2)</sup> (s)
<b>conf1</b>	6112 (2074, 1970, 2068)	18	1076	3.071	20.24
<b>conf2</b>	6227 (2509, 1845, 1873)	13	1405	3.810	22.01
<b>conf3</b>	7219 (3321, 2455, 1443)	10	3289	4.990	36.18
<b>conf4</b>	6084 (2584, 1916, 1584)	11	1889	3.478	23.97
<b>conf5</b>	6177 (1934, 2548, 1695)	15	1088	3.955	19.09
<b>conf6</b>	7206 (3178, 2540, 1488)	13	2500	4.162	27.95
<b>conf7</b>	5462 (2093, 1828, 1541)	11	666	3.018	16.03
<b>conf8</b>	6350 (2806, 2172, 1372)	10	2139	2.735	25.91
<b>conf9</b>	5765 (2178, 2116, 1471)	12	2081	3.303	30.96
<b>conf10</b>	5240 (1614, 2025, 1601)	11	211	4.084	8.82
<b>conf11</b>	5700 (1664, 2251, 1785)	15	273	4.603	9.97
<b>conf12</b>	5528 (1771, 2302, 1455)	14	356	3.711	11.45
<b>conf13</b>	5426 (1462, 2233, 1731)	10	190	4.409	8.11
<b>conf14</b>	6949 (2468, 2610, 1871)	19	2839	3.756	37.18
<b>conf15</b>	5907 (2031, 2207, 1669)	11	1195	3.792	21.14
<b>conf16</b>	6524 (2476, 2288, 1760)	12	2455	4.110	32.63
<b>conf17</b>	5436 (2017, 1813, 1606)	20	1287	1.909	19.50
<b>conf18</b>	6047 (2084, 2349, 1614)	13	819	3.842	17.12
<b>conf19</b>	6151 (2134, 1979, 2038)	18	1261	2.590	21.90
<b>conf20</b>	6095 (2310, 2301, 1484)	11	1566	3.584	22.13

1) Anzahl der Domänenpaare mit einer Komplementarität über 0.8

Die erste Zahl ist die Gesamtzahl aller Domänenpaare, in Klammern sind die nach topographischen Kriterien, nach dem elektrostatischen Potential und nach der Lipophilie eingeteilten Domänen getrennt aufgeführt.

Für die Berechnung des elektrostatischen Potentials wurde der modifizierte Coulomb-Ansatz verwendet.

2) Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor

Tabelle A6.5: Ergebnisse des Tests auf Volumenüberlappung und des Clusteralgorithmus (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	Anzahl der Cluster	<i>rms</i> -Wert (Å) des besten Clusters	Rechenzeit (s)
conf1	270	3.513	1.33
conf2	271	3.810	2.75
conf3	514	5.225	7.43
conf4	316	3.487	6.43
conf5	299	3.969	1.67
conf6	470	4.162	5.22
conf7	175	3.474	0.60
conf8	325	2.735	4.00
conf9	333	3.757	3.75
conf10	92	4.064	0.10
conf11	102	4.736	0.16
conf12	148	3.711	0.24
conf13	81	4.474	0.13
conf14	519	4.045	7.17
conf15	307	3.810	2.37
conf16	371	4.184	6.58
conf17	236	2.485	1.98
conf18	259	4.052	1.04
conf19	269	2.903	2.56
conf20	308	3.689	2.24

Tabelle A6.6: Ergebnisse der rigiden Energieminimierung mit dem *Downhill Simplex*-Algorithmus (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	<i>rms</i> -Wert (Å) der besten Struktur	Rang der besten Struktur	Energie der besten Struktur	Rechenzeit (s)
conf1	3.554	235	-17	26.28
conf2	5.181	25	-67	26.16
conf3	5.510	353	-35	42.06
conf4	4.920	307	-1	29.82
conf5	4.066	192	-37	28.05
conf6	5.804	262	-42	39.51
conf7	4.142	160	4	20.08
conf8	2.598	291	-22	29.99
conf9	2.920	18	-65	30.68
conf10	4.542	32	-56	14.53
conf11	5.207	42	-48	15.00
conf12	4.444	22	-68	17.96
conf13	3.354	8	-71	13.66
conf14	3.699	84	-61	42.79
conf15	3.367	79	-55	28.73
conf16	4.113	328	-11	32.89
conf17	1.203	4	-83	24.22
conf18	3.257	75	-59	25.63
conf19	3.275	45	-58	26.32
conf20	3.896	290	-7	28.94

Tabelle A6.7: Ergebnisse des *Clustering* der minimierten Komplexstrukturen (modifizierter Coulomb-Ansatz). Die Rechenzeit bezieht sich auf eine Intel-basierte Linux-Workstation mit 500 MHz Pentium III Prozessor.

<i>pdb</i> -Datei	Anzahl der Cluster	<i>rms</i> -Wert (Å) des besten Clusters	Rechenzeit (s)
conf1	182	3.692	0.20
conf2	181	5.181	0.22
conf3	340	5.510	0.46
conf4	205	4.920	0.30
conf5	193	4.285	0.23
conf6	301	5.815	0.38
conf7	124	4.070	0.11
conf8	234	2.601	0.24
conf9	246	3.277	0.24
conf10	79	4.916	0.05
conf11	77	5.282	0.05
conf12	113	4.444	0.10
conf13	64	3.354	0.04
conf14	334	3.700	0.42
conf15	205	3.367	0.27
conf16	241	4.113	0.32
conf17	176	1.205	0.16
conf18	171	3.282	0.13
conf19	168	3.267	0.21
conf20	212	3.896	0.19

## Hilfsmittel

Teile der Auswertung wurden mit den Molecular Modeling Programmpaketen *SYBYL* und *CHARMM*, die Visualisierung der Molekülgraphiken mit dem Programm *MOLCAD II* sowie ISIS Draw von MDL Information Systems, Inc. erstellt. Für die Entwicklung des genetischen Algorithmus wurden ProgrammROUTINEN der *Genetic Algorithm Library* von Levine verwendet. Zur Erstellung der schematischen Darstellungen und der mathematischen Funktionen dienten die Programme Paint Shop Pro 6.0 der Firma Jasc Software, Micrografx Designer 6.0, Micrografx ABC FlowCharter 6.0 sowie Origin 5.0 der Firma Microcal. Der vorliegende Text wurde mit Microsoft WORD 97 geschrieben.



**Lebenslauf:**

Name: Thomas Eckart Exner

Geburtsdatum: 26. Juni 1972

Geburtsort: Stuttgart

Staatsangehörigkeit: österreichisch

Schulbesuch: 1982-1983  
Gerhart-Hauptmann-Realschule, Leonberg

nach der Fertigstellung des Schulneubaus

1983-1988

Ostertag-Realschule, Leonberg

1988-1991

Technisches Gymnasium, Berufliches Schulzentrum Leonberg

Studium: 1991-1996

Studium der Chemie an der Technischen Hochschule Darmstadt

Abschluß: Diplom

Promotion: 1996-2000

Promotion im Arbeitskreis von Prof. Dr. J. Brickmann,

Institut für Physikalische Chemie I

Technische Universität Darmstadt

**Eidesstattliche Erklärung**

Hiermit erkläre ich an Eides statt, daß ich meine Dissertation selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Darmstadt, Januar 2000